THE PURINE WORLD: EXPERIMENTAL INVESTIGATIONS INTO THE PREBIOTIC SYNTHESIS OF PURINE NUCLEOBASES AND INTERCALATION OF HOMOPURINE DNA DUPLEXES

A Dissertation
Presented to
The Academic Faculty

by

Ragan Buckley

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology

August 2012

THE PURINE WORLD: EXPERIMENTAL INVESTIGATIONS INTO THE PREBIOTIC SYNTHESIS OF PURINE NUCLEOBASES AND INTERCALATION OF HOMOPURINE DNA DUPLEXES

Approved by:

Dr. Nicholas V. Hud, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Thomas M. Orlando
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Jean-Marie D. Dimandja
School of Chemistry
Spelman College

Dr. Loren D. Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Sheldon W. May
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: May 24, 2012

It is often said that all the conditions for the first production of a living organism are now present which could ever have been present. But if (and oh what a big if) we could conceive in some warm little pond with all sorts of ammonia and phosphoric salts, light, heat, electricity etc. present, that a protein compound was chemically formed, ready to undergo still more complex changes, at the present day such matter would be instantly devoured, or absorbed, which would not have been the case before living creatures were formed.

Charles Darwin, Letter to Joseph D. Hooker (1871)

ACKNOWLEDGMENTS

I started graduate school with no prior research experience and only a vague idea of what would be required of me. I think – hopefully rightly so – that I've gained valuable skills and a deeper understanding of scientific inquiry along the way. I couldn't have done it alone, though.

First and foremost, I'd like to thank my advisor, Nick Hud. In his lab, the perfect balance was struck between watchful guidance and individual initiative. I learned to function on my own, as a scientist, but always knew help was available when I needed it – just a knock on the office door (or a Skype call) away. Nick's attitude that we should understand the workings of our own equipment and instruments has led me to performing all manner of tasks that I would never have expected to be doing – taking apart vacuum pumps, replacing seals on an HPLC, wiring a temperature controller, and more – and I am grateful for those experiences.

The rest of my Ph.D. committee is also deserving of thanks. Thom Orlando has been like a second advisor to me, especially with some of my later work with iron-containing minerals. I share John Dimandja's interest in and enthusiasm for chromatography of complex mixtures and thank him for his lucid explanations of two-dimensional gas chromatography and also for a supply of excellent undergraduate research assistants. Loren Williams has influenced me to think more deeply about both the thermodynamics of DNA intercalation and the actual conditions present on the early Earth. And Sheldon May has provided a valuable outside perspective; he asks the important questions that don't always occur to specialists in prebiotic chemistry.

My graduate education would not have been the same without the insights and experiences gained through my interactions with my labmates. I am grateful to Heather Bean, Özgül Persil Çetinkol, Tumpa Sarkar, Eric Horowitz, Aaron Engelhart, Irena Mamajanov, Denise Okafor, Tim Lenz, Brian Cafferty, Brandon Laughlin, Ford Lannan, Seth Lilavivat, Shreyas Athavale, Sara Imari Walker, Samanta Pino, and Michael Chen for their help and advice. I have been lucky to have supervised a number of excellent undergraduates in the lab; I'd like to thank Christina Graves, Whitney Smith, Shaquitta Bell, Amanda Bryant, Scott Hoy, Amena Warrayat, Michael Ruemmele, and Jaheda Khanam for their research contributions. Jaheda deserves additional thanks for all her help readying supplies for high school lab demos; I never would have gotten all of those samples prepared without her.

All of my published work was enriched by contributions from co-authors and collaborators. Many of these individuals have already been named above, but I wanted to give credit to those who have not yet been mentioned. Hannah Barks and I worked closely together on the first formamide paper and Greg Grieves was instrumental in helping with the design and improvement of the formamide reaction centers. Ernesto Di Mauro provided valuable feedback and advice with respect to formamide chemistry. Lakshmi Anumukonda and Avery Young helped me to understand both the challenges and the rewards of involvement with chemical education at a pre-college level, and David Lynn's suggestions for the high school lab project dramatically improved the *Journal of Chemical Education* manuscript.

Several members of the Center for Chemical Evolution have also provided valuable feedback on my experiments and results. I've gained a lot from my interactions

with other students working on formamide projects, including Rachel Bennett and Michele Dawley. I also appreciate discussions with Jeff Bada regarding the prebiotic environment and I'm thankful to Greg Springsteen for his comments on formamide chemistry and for providing me with unpublished results that were directly relevant to my own work.

There were times when I had to look outside the Hud lab for assistance with experiments, equipment, or training; I'm indebted to a number of individuals for their help with these aspects of my research, as well. The Schuster group responded promptly and amiably to my repeated requests for DNA synthesis. Nadia Boguslavsky was extremely patient and helpful in answering my beginner's HPLC questions during my first year at Georgia Tech. David Bostwick and others at the Bioanalytical Mass Spectrometry Facility have been invaluable with respect to mass spec training, instrument access, and troubleshooting. Though NMR spectroscopy did not figure prominently in the final versions of my projects, I am grateful to Les Gelbaum for instruction in the use of NMR instruments and for useful discussions and suggestions with respect to my NMR experiments. I'd like to thank Facundo Fernandez for the use of an Agilent 1100 HPLC; having this instrument to myself was vital to making progress on my iron mineral work. I'd also like to thank Leigh Bottomley and Richard Bedell. Leigh provided me with a portable UV-Vis spectrometer and Richard constructed a custom cable to connect the two halves of this spectrometer. Without them, I never would have been able to complete my dilution experiments. I'm also grateful to Kyril Solntsev for his assistance with my fluorescence experiments and to Tony Brand at Agilent for an automated HPLC data export macro.

A number of other people have offered other vital assistance during my years at Georgia Tech, and I'd be remiss if I neglected to mention them. Sue Winters, Gary Longstreet, Christine Conwell, Jamila Cola, Meisa Salaita, Cam Tyson, and all the IBB support staff – thanks to you, as well!

Finally, I'd like to acknowledge Patrick Yaner's willingness to discuss obscure usage concerns and unusual grammatical constructions. I'm sure he rolled his eyes every time I told him I was going to be working from home on this document – he knew that meant I'd be sending e-mails with language questions that simply didn't have straightforward or easy answers.

If I've neglected to mention anyone, the omission is entirely my own and I profusely apologize.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

7	7-deazaxanthine/7-deazaxanthosine
9	isoguanine/isoguanosine
A	adenine/adenosine
Å	angstrom
AICA	aminoimidazole carboxamide
AICN	aminoimidazole carbonitrile
arb. units	arbitrary units
ATP	adenosine triphosphate
AU	absorbance units
aza3	azacyanine3
B-DNA	B-form DNA
B3LYP/6-31G*	a density functional theory model and basis set
bp	base pair
°C	degrees Celsius
C	cytosine/cytidine
C8	octylsilane
C18, ODS	octadecylsilane
CD	circular dichroism
cm	centimeter
D	dextrorotatory
D	2,6-diaminopurine (base or nucleoside)

D_0	total number of ligand molecules in the system
Da	Dalton
dA	deoxyadenosine
DAD	diode array detector
DAFN	diaminofumaronitrile
DAMN	diaminomaleonitrile
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dT	deoxythymidine
ΔT_m	change in melting temperature
ε	extinction coefficient, molar absorptivity coefficient
EDTA	ethylenediaminetetraacetic acid
eV	electron volt
FAD	flavin adenine dinucleotide
fNA	flexible nucleic acid
fs	femtoseconds (10^{-15} seconds)
g	grams
G	guanine/guanosine
Ga	billion years
GC-MS	gas chromatography-mass spectrometry
gly-NA	glyoxylate nucleic acid
GNA	glycol nucleic acid
GTP	guanosine triphosphate

h	hours
^{1}H	proton (in NMR)
HCN	hydrogen cyanide
HNC	hydrogen isocyanide
HPLC	high pressure liquid chromatography
$h\nu$	energy (Planck's constant × frequency)
I	inosine (hypoxanthine nucleoside)
IDP	interplanetary dust particle
K	Kelvins (temperature)
K_a	equilibrium association constant
kcal	kilocalorie
λ	wavelength
L	levorotatory
L	liter
LC	liquid chromatography
LC-MS	liquid chromatography-mass spectrometry
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LHB	late heavy bombardment
LUCA	last universal common ancestor
M^+	metal ion
MD	molecular dynamics
mg	milligram
MHz	megahertz

μ, μm	micron (10^{-6} meters)
μL	microliter (10^{-6} liters)
mL	milliliter
mm	millimeter
mM	millimolar
mol	moles (amount of substance)
mol/mol	mole percent
MRM	multiple reaction monitoring
mRNA	messenger RNA
MS/MS	tandem mass spectrometry
m/z	mass-to-charge ratio
NAD+	nicotinamide adenine dinucleotide (oxidized form)
NH_4OAc	ammonium acetate
nm	nanometer (10^{-9} meters)
NMM	*N*-methyl mesoporphyrin
NMR	nuclear magnetic resonance
NPT	constant number of atoms, pressure, and temperature
ns	nanoseconds (10^{-9} seconds)
p	phosphate linkage
PDB	Protein Data Bank
PEEK	polyether ether ketone
PEP	phosphoenolpyruvate
PFP	pentafluorophenyl

pH	potential hydrogen
PID	proportional-integral-derivative
pK_a	acid dissociation constant
PME	particle mesh Ewald
PNA	peptide nucleic acid
ppm	parts per million
ps	picoseconds (10^{-12} seconds)
Pu·Pu	purine-purine (refers to specific DNA strand)
R	generic sugar unit (in chemical structure); ratio (in Job plot)
(*R*)	rectus (right-handed)
RNA	ribonucleic acid
RRLC	rapid resolution liquid chromatography
s	seconds
(*S*)	sinister (left-handed)
S_0	total number of binding sites
SHAKE	algorithm for satisfying bond geometry constraints
SIC	selected ion chromatogram
super.	supernatant
T	thymine/thymidine
TIC	total ion chromatogram
TIP3P	transferable intermolecular potential three point
TLC	thin layer chromatography
T_m	melting temperature

TMPyP4	tetrakis(4*N*-methylpyridyl)porphine
TMS	tetramethylsilane
TNA	threose nucleic acid
TPD	temperature programmed desorption
Tris	2-amino-2-hydroxymethylpropane-1,3-diol
tRNA	transfer RNA
UV	ultraviolet
UV-Vis	ultraviolet-visible
V	volt
v/v	volume percent
w/v	weight/volume percent
WC	Watson-Crick (refers to specific DNA strand)
x	concentration of ligand bound
X	xanthine/xanthosine
ZIC-HILIC	zwitterionic hydrophilic interaction chromatography

SUMMARY

The origin of the first biopolymers from abiotic precursors has generated considerable interest in the scientific community. While contemporary nucleic acids possess superior functionalities, DNA and RNA were likely preceded by a different polymer or set of polymers which may have been easier to synthesize and assemble. The recognition elements of these proto-RNAs, as well as their sugars and backbone linkages, likely differed to some degree from the canonical bases (adenine, guanine, thymine/uracil, and cytosine), from the ribose or deoxyribose sugar, and from the phosphate linkage observed today.

Formamide is the simplest molecule containing the four essential elements of life: carbon, hydrogen, nitrogen, and oxygen; it has been detected on comets and in the interstellar medium and is a known hydrolysis product of hydrogen cyanide (HCN). Formamide is a solvent of great interest to prebiotic chemists because it is liquid over a wide range, it is less volatile than either water or HCN, and it possesses a versatile reactivity. When formamide is heated in the presence of minerals or inorganic catalysts, a variety of products including purine nucleobases are generated. Irradiation of formamide reaction solutions with ultraviolet light increases the yield and diversity of products, and eliminates the need for a mineral catalyst.

When designing model prebiotic reactions, it is important to consider the geology and atmosphere of the early Earth. To date, most formamide reactions have been conducted under an air atmosphere and catalyzed by a range of minerals and salts, some of which may not be prebiotically relevant. We have performed formamide reactions in

the presence of pyrite, an iron sulfide mineral which is likely to have been available on the primordial Earth. Further, we have investigated the effects of surface reactions versus solution phase reactions, and we have conducted formamide experiments under different atmospheres to better gauge the effects of gases on the products generated during these reactions. Our results indicate the greatest yield and diversity of products result from the combination of a pyrite mineral catalyst, heat, UV irradiation, and a carbon dioxide atmosphere.

Phosphate minerals are important, biologically, and must have been incorporated into primordial biopolymers at an early stage of chemical evolution. Further, iron in its Fe^{2+} form is believed to have been widely available at the time life originated. Iron phosphate minerals are thus of great interest to prebiotic chemists. Additionally, it is unlikely that a pool of pure formamide ever existed on the early Earth; a diverse set of one-, two-, and three-carbon molecules may also have been present. We have performed formamide experiments in the presence of various iron phosphate minerals and, separately, with the addition of small, carbon-containing molecules, to examine the catalytic effects of these minerals on and the contributions of these small molecules to the product profiles of our reactions.

Purine nucleobases are simple to synthesize in model reactions and they stack well in aqueous solution. The formation of antiparallel, homopurine DNA duplexes has recently been demonstrated. It has been hypothesized that the first nucleic acids were composed of only purine bases, and that water-soluble, cationic, aromatic molecules with large stacking surfaces ("molecular midwives") may have aided the assembly of the earliest nucleic acid analogs. We have characterized the interactions of various

intercalators with a standard DNA duplex as well as with an antiparallel homopurine DNA duplex and have determined that molecules which possess four or more rings and a curved shape interact selectively with all-purine DNA; such molecules can serve as models for putative prebiotic midwives.

We have also demonstrated the production of large heterocyclic molecules in heated formamide solutions. While these minor reaction products remain to be fully characterized, a one-pot synthesis of both nucleobase and midwife molecules is attractive from the standpoint of parsimony in chemical evolution.

CHAPTER 1

INTRODUCTION

1.1 THE ORIGIN OF LIFE ON EARTH

The first organisms on Earth were unicellular prokaryotes or protoeukaryotes and their existence has been dated as far back as 3.465 to 3.4 billion years (Ga), though identification and description of the earliest microfossils remains controversial [2-4]. While the period of planetary accretion ended between 4.56 and 4.45 Ga [5], the late heavy bombardment (LHB) of 3.95 to 3.8 Ga, to which the resurfacing of the entire planet is attributed, likely created conditions unfavorable for the survival of life's molecules [5]. Thus, the origin of extant life on Earth is currently believed to have occurred between approximately 3.8 and 3.465 Ga. Life may well have arisen multiple times and in multiple locations around the globe during this period, though today, all contemporary life forms are thought to have descended from a single organism known as the Last Universal Common Ancestor (LUCA) or, more probably, from a community of such organisms engaged in horizontal gene transfer [4].

Life has been defined in numerous ways [6], though it is generally agreed that all modern organisms adhere to the Central Dogma of Molecular Biology. (Viruses, while inextricably tied to life, lack the capability to replicate themselves without appropriating the machinery of host cells.) The Central Dogma, first proposed by Francis Crick, states that information transfer in biology – with a few exceptions – proceeds unidirectionally from deoxyribonucleic acid (DNA) to ribonucleic acid (RNA) to protein [7].

It is improbable, however, that the LUCA organism – or community of organisms – sprang, fully-formed, from the prebiotic environment. Various encapsulated pre-cellular systems have been proposed, including proteinoid microspheres [8-10] and lipid vesicles and micelles [11-15], though it has thus far been impossible to experimentally duplicate the transition from one of these models to a fully functioning, self-maintaining, living system [16].

1.2 THE RNA WORLD HYPOTHESIS

The LUCA organisms are believed to have adhered to the Central Dogma, using DNA for information storage, RNA for information transfer, and proteins for catalysis. The interplay of these three biological macromolecules, each of which is uniquely situated to its purpose in the cell, presupposes a complex evolutionary history. The question of which of these molecules – DNA, RNA, or protein – originated first was answered through a series of experiments conducted beginning in the 1950s and continuing through the 1980s and beyond.

The work of Erwin Chargaff on the base composition of DNA demonstrated that the percentage of adenine in an organism's DNA is equivalent to the percentage of thymine, and that the amounts of guanine and cytosine also matched one another [17-19]. In 1953, Watson and Crick discovered that DNA forms a double helix [20, 21]. Together, the results of these experiments strongly suggested a mechanism of information transfer whereby the base sequence of one strand was copied into a complementary base sequence in the paired strand in a semi-conservative manner; this was confirmed by Meselson and Stahl via centrifugation experiments in 1958 [22].

Because DNA was firmly ensconced as the carrier of genetic information, it was surprising to many researchers that two complementary RNA molecules (polyadenylic acid and polyuridylic acid) were also able to form a double helical structure [23]. RNA (Figure 1.1) is similar to DNA in that it possesses a backbone of alternating sugar and phosphate moieties and four nitrogenous bases that act as recognition elements.

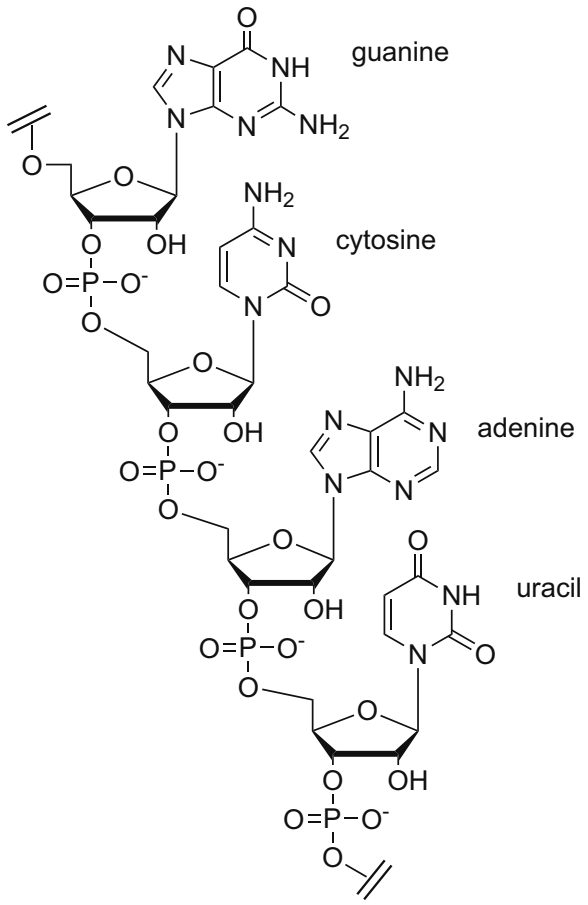

Figure 1.1 Chemical structure of RNA.

The work of Rich and Davies made it apparent as early as 1956 that RNA was also capable of storing information [23]. As time went on, other scientists speculated as to the catalytic nature of RNA [24, 25], but it was not until the early 1980s that the Cech

and Altman labs, working independently, discovered that RNA also possessed the catalytic properties of self-splicing to remove introns [26] and cleavage of tRNA precursor molecules [27], respectively. These findings led Walter Gilbert to coin the term "RNA World" to describe a pre-life scenario wherein RNA performed the functions associated today with DNA and proteins [28]. As proteins and DNA later evolved, these molecules took over catalysis and information storage functions. However, some vestiges of the RNA World may remain in contemporary organisms; for example, the peptidyl transferase center in the ribosome is a ribozyme, or RNA enzyme [29-32].

1.3 THE PRE-RNA WORLD

Although RNA possesses both information storage and catalytic properties, it is a complex, chirally-pure polymer [33-35] with monomer units linked together by high-energy phosphodiester bonds. RNA is thus likely to have been preceded by a nucleic-acid-like polymer which would have been easy to assemble from component parts available on the prebiotic Earth, but with an inferior functionality inferior.

Most RNA World scenarios assume Watson-Crick-type base-pairing as the means by which successful sequences could replicate [34], and many different purine and pyrimidine nucleobases (Figure 1.2) have been generated in model prebiotic reactions [36-40]. The nucleobases in primordial informational polymers need not have been the same as those present contemporary nucleic acids; multiple alternative purine-pyrimidine [41, 42] and purine-purine [24, 43-45] base pairs have been postulated.

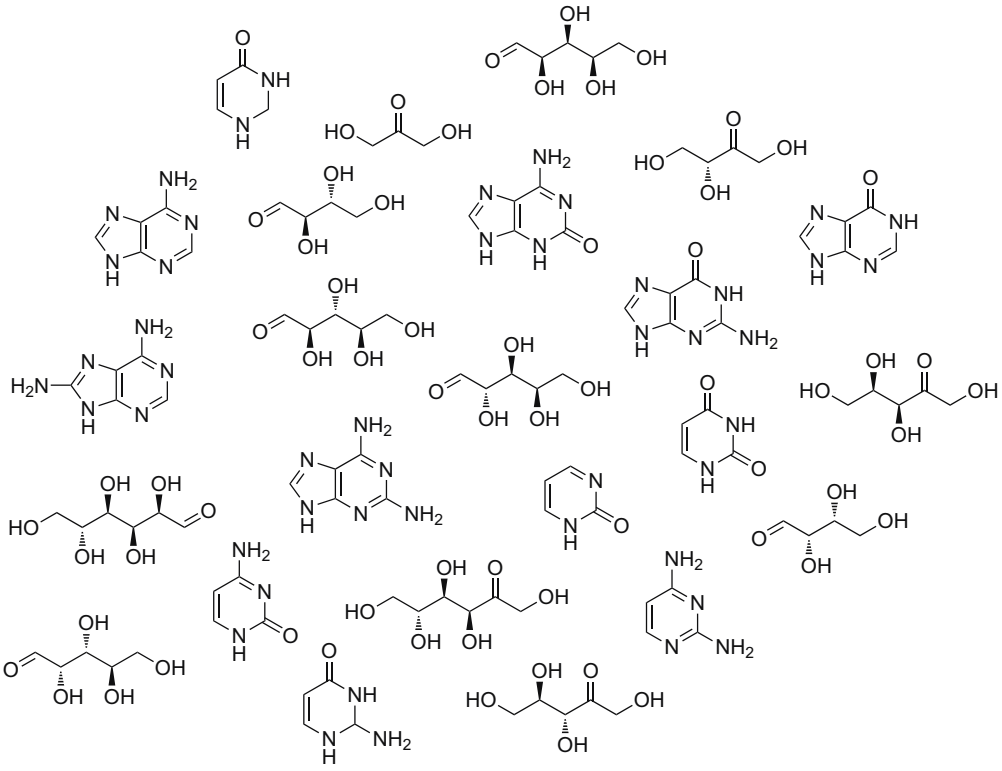

Figure 1.2 A selection of nucleobases and sugars which may have been present in the first nucleic acids. Pentose and hexose sugars, shown here in their linear forms, are able to cyclize into furanose (five-membered) and pyranose (six-membered) rings in aqueous solutions. For all sugars, both L and D forms were likely present, though only D forms are pictured above.

Various sugars (Figure 1.2) have also been produced in prebiotic reactions [46-48], and functional nucleic acid enzymes with a mixture of sugar units have recently been described [49]. It is conceivable that the first nucleic acids incorporated a variety of canonical and non-canonical nucleobases and multiplicity of sugar units with both L and D chiralities. Backbone linkage variation may also have occurred. Over time, the primitive genetic polymers of the pre-RNA world would have been subjected to selective pressures, eventually evolving into RNA.

1.4 THE PREBIOTIC MILIEU

Some five or six decades after Darwin first speculated about his "warm little pond" in a letter to Joseph Hooker [1], Russian biochemist Alexander I. Oparin proposed that the atmosphere of the primitive Earth would have been reducing, consisting of methane, ammonia, hydrogen gas, and water vapor; his original work was not widely disseminated and a few years later, British geneticist J.B.S. Haldane put forward a similar idea, substituting carbon dioxide for Oparin's methane [50-54]. Oparin and Haldane further hypothesized that energy sources such as lightning or ultraviolet radiation would have allowed the production of small organic molecules from inorganic precursors (the "heterotrophic origin-of-life theory") [50-54]. While many scientists today believe the atmosphere of the early Earth was not as reducing as originally postulated, the question is far from settled, and the Oparin-Haldane hypothesis has provided a framework for much of the research in prebiotic chemistry that has occurred over the last sixty years.

1.4.1 The Prebiotic Atmosphere

Early efforts to synthesize biomolecules from simple, inorganic components relied on a reducing atmosphere modeled after Oparin's hypothesis [55, 56]. It is now nearly universally accepted that the planetary atmosphere contained no oxygen until the advent of photosynthetic microorganisms [57], though both opinions and experimental evidence differ with respect to which gases were present.

According to some current hypotheses, the early Earth's atmosphere was rich in carbon dioxide [58-60]. Other suggestions from the last decade are based on various lines of geological and chemical reasoning and include hazes composed of nitrogen and methane with a significant ammonia component [61, 62], carbon dioxide with a large

percentage of molecular hydrogen [63], and carbon monoxide-dominant atmospheres [64].

1.4.2 Ultraviolet Radiation

While the Oparin-Haldane hypothesis posited ultraviolet radiation in the absence of an ozone-containing atmosphere as one possible energy source for the synthesis of simple organic compounds [50, 51], scientists working in the 1960s and later expressed concern that radiation could be damaging to nascent life [65-68]. Various protective mechanisms have also been suggested that may alleviate some of these concerns; among these are submersion of early life forms in the deep ocean [69] and shielding by dissolved ferric iron [70].

Prior to the advent of the first cells, however, UV light may have actually assisted in the formation of organic molecules. UV-aided syntheses of nucleobases [40, 71, 72] have been reported, and ultraviolet radiation has been demonstrated to assist in the generation of sugar precursors from a variety of atmospheric gases [73-77].

1.4.3 Minerals and Ions

Mineral surfaces, including clays, have long been of interest to the prebiotic research community. These surfaces have been invoked in every role from catalysts promoting the formation of nucleobases [37, 38, 78-80], to sources of phosphate for phosphorylation of nucleosides [81-84], to substrates that aid in the polymerization of activated nucleotides [85, 86].

Iron-containing minerals have generated particular interest; this group includes iron sulfides such as pyrite, pyrrhotite, and marcasite, as well as the iron nickel phosphide schreibersite. Iron sulfides are thought to be particularly ancient, dating as far back as the

end of the era of Earth's accretion [87]. They have long been invoked by researchers and theoreticians as sources of energy for primitive metabolic cycles [88-91], although the feasibility of such cycles has recently been called into question [92]. Schreibersite and similar minerals, meanwhile, are typically found in meteorites and may have proved valuable as sources of reactive phosphorus during the early stages of chemical evolution [93]. Iron and other transition metals are characterized by rich coordination and oxidation-reduction chemistries which could have been vitally important to the synthesis of biomolecules from simple inorganic precursors.

Iron sulfides such as pyrite are insoluble, despite their iron being present in a reduced state [94]. It is widely believed, however, that the oceans of the early Earth contained concentrations of ferrous (Fe^{2+}) iron as high as several hundred micromoles per liter [95, 96]. Weathering, volcanism, and circulation of water from near the sea floor have all been invoked to explain the presence of iron in Hadean oceans [96]. Regardless of the ultimate source, dissolved iron must be considered alongside mineral-bound iron in any discussion of prebiotic catalysis; the same is true for other metals of interest.

1.4.4 The Temperature of the Early Earth

Various temperatures have been proposed for the Earth at the time of life's origin. The "faint young sun" paradox stems from estimates that the early Sun was approximately thirty percent less luminous at the time life is hypothesized to have arisen on Earth; this decreased radiative intensity would have resulted in temperatures on the planetary surface which are incompatible with the existence of liquid water [97-100]. Liquid water's importance to life today is unquestioned, and various ideas have been proposed to bypass the faint young sun issue.

A cold origin of life is one answer to the paradox. It has been noted that hydrogen cyanide (HCN) is unlikely to have polymerized in a warm (e.g. 100°C) ocean, whereas eutectic freezing of water can favor HCN polymerization [101]. Several studies have reported the synthesis of nucleobases in frozen solutions of ammonium cyanide [102, 103]; others have employed freeze-thaw cycles of aqueous solutions of urea to obtain similar compounds [104], simulating the effects of localized thawing due to meteoritic strikes [105]. Still others have proposed that high concentrations of carbon dioxide [58-60] or other greenhouse gases [62] would have raised the temperature of the Earth's surface, despite the existence of a faint young sun. Scientists have also invoked a lower planetary albedo as a way to moderate surface temperatures [106].

1.4.5 The Chemistry of HCN, Formamide, and Ammonium Formate

Hydrogen cyanide and its hydrolysis product, formamide, are of considerable importance in origin-of-life research. The earliest reported syntheses of adenine from simple inorganic precursors involved HCN polymerizations [39, 107]; in fact, adenine is sometimes still referred to as the "HCN pentamer" because of its chemical formula ($C_5H_5N_5$). HCN is ubiquitous in the universe and HCN polymer is believed to be a major constituent of tholins [108-112], which make up the nitrogen-rich organic hazes surrounding comets and icy moons. Ongoing research attempts to more fully elucidate the structure and properties of HCN polymers [113-115]. Ammonium formate, the hydrolysis product of formamide, also contributes to the synthesis of adenine from the precursor diaminomaleonitrile, or DAMN [116].

1.4.5.1 Chemical Equilibria

Chemists wishing to use formamide as a solvent for electrochemical processes require a high purity, and efforts to remove water from formamide have proven uniquely challenging. Many procedures which are employed in the purification other solvents, such as the use of cation and anion exchange resins, can actually promote the hydrolysis of formamide to ammonium formate [117]. Heating of formamide solutions also increases the rate of hydrolysis (Figure 1.3). At ambient temperature, the half-life of this process is roughly two hundred years [36], but at 130°C, approximately one percent of formamide is converted to ammonium formate in twenty-four hours [40].

Formamide can also decompose into HCN and water (Figure 1.3); this reversible transition is enhanced by heat, certain mineral catalysts, and a solution pH which is slightly basic [36]. Since formamide and water do not form an azeotrope [118], during prolonged heating, water will evaporate from mixed water/formamide solutions [40], thereby adding further complexity to the composition of formamide solutions.

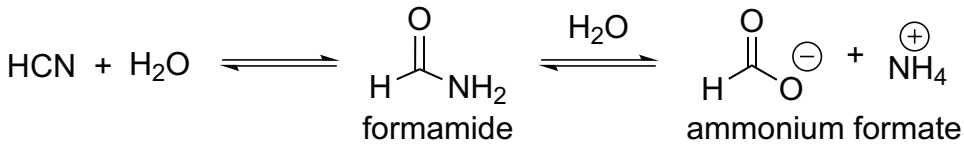

Figure 1.3 Chemical equilibrium between hydrogen cyanide (HCN), formamide, and ammonium formate.

In any likely prebiotic scenario involving HCN, water, formamide, and ammonium formate, a mixture of reactive species is certainly present. While this presents a problem for precise electrochemical work, the diversity of available chemical

groups in this system – nitrile, amine/amide, and carboxylate – is a boon for the origin-of-life chemist seeking to maximize product diversity in model reactions.

1.4.5.2 Synthesis of Adenine from HCN

Adenine is omnipresent in contemporary organisms – it appears in DNA, RNA, the ATP (adenosine triphosphate) that powers much of a cell's metabolism, and the redox cofactors FAD (flavin adenine dinucleotide) and NAD+ (nicotinamide adenine dinucleotide). It is also ubiquitous in prebiotic chemistry, having been identified in meteorites [119] and as a product of formamide reactions [36, 40, 80]. Perhaps the best-known synthesis of adenine, however, starts from HCN (Figure 1.4).

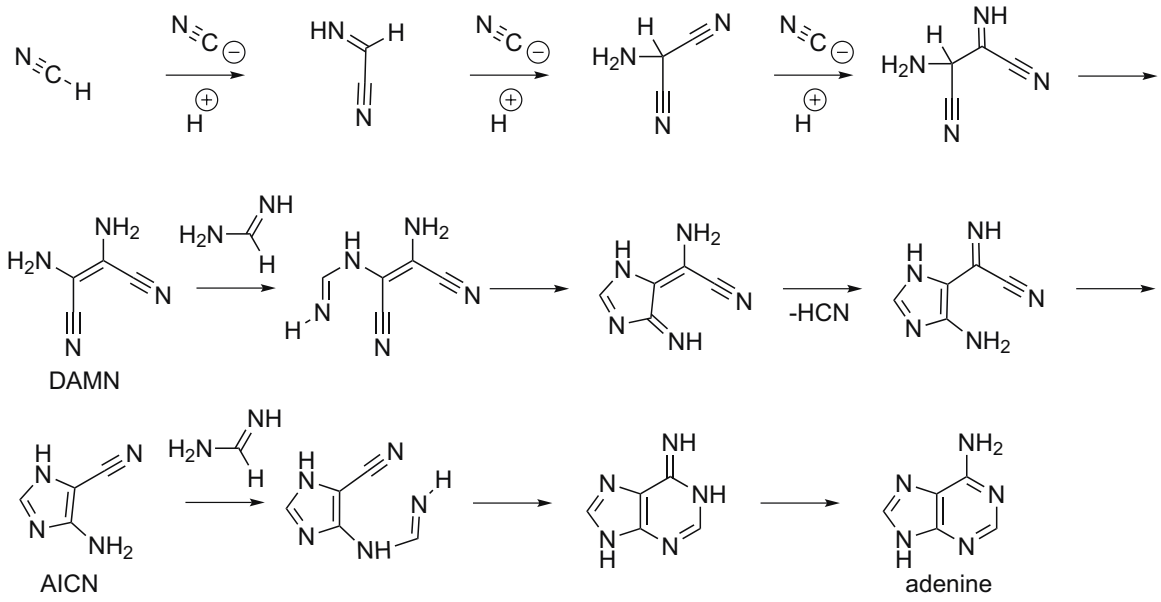

Figure 1.4 Proposed pathway of adenine synthesis from HCN [39, 107, 120-124].

Adenine is unique among nucleobases in that it was likely one of the earliest purines generated and in that it has remained essential to contemporary life. Francis

Crick hypothesized that adenine was a vital component of the first informational polymers and noted that its hydrolysis product, inosine (the hypoxanthine nucleoside), may have formed base pairs with adenine in these polymers [24].

1.4.5.3 Prebiotic Syntheses from Formamide

The synthesis of purine from formamide (Figure 1.5) has been known for some time [125, 126], but formamide chemistry has received a great deal of additional attention in the early part of the twenty-first century [36-38, 40, 79, 80, 127, 128].

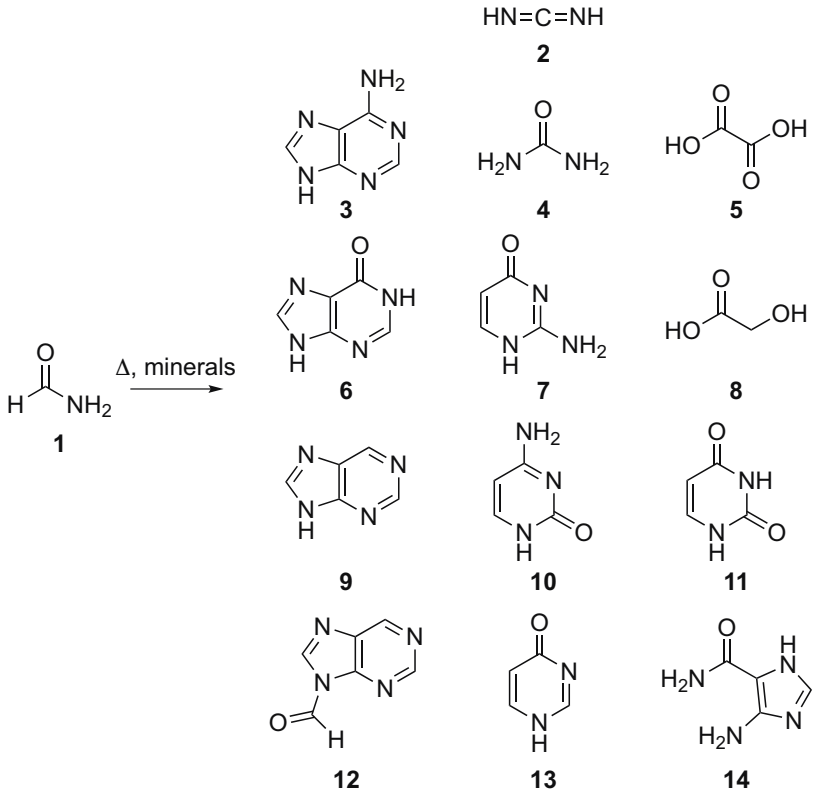

Figure 1.5 Selected formamide reaction products. **1** formamide, **2** carbodiimide, **3** adenine, **4** urea, **5** oxalic acid, **6** hypoxanthine, **7** isocytosine, **8** glycolic acid, **9** purine, **10** cytosine, **11** uracil, **12** 9*H*-formylpurine, **13** 4(6)-hydroxypyrimidine, **14** AICA (aminoimidazolecarboxamide).

Simple heating of formamide at 160°C in the presence of various minerals, clays, and simulated [129] or actual [128] extraterrestrial material leads to the production of a variety of products with prebiotic relevance, including nucleobases, substituted imidazoles, acyclonucleosides, carboxylic acids, and small molecules such as urea and carbodiimide (Figure 1.5). In the absence of ultraviolet irradiation, the identity of the mineral does appear to affect the outcomes of reactions in terms of both product diversity and yield [36], though when formamide reactions are irradiated with 254 nm UV light, the characteristics of the mineral or salt present appear to matter less [40]. Thus far, no one studying this chemistry has investigated the role of ions in solution versus the role of species trapped in a mineral matrix.

1.4.6 Other Noted Prebiotic Syntheses

While much formamide and HCN chemistry has focused on the production of nucleobases, life also makes use of other simple monomers such as sugars and amino acids, and syntheses of these molecules have also been reported.

1.4.6.1 The Miller-Urey Experiment

Perhaps the most iconic experiment in prebiotic chemistry was the Miller-Urey experiment. In 1953, University of Chicago graduate student Stanley Miller constructed an apparatus that subjected methane, ammonia, and hydrogen to an electric discharge. Heated water simulated an ocean and contributed water vapor to the gas mixture; a condenser allowed water and other substances to collect and exit the vapor phase. Miller allowed the apparatus to run for several days – a color change occurred on the first day of the experiment – and then analyzed the products by paper chromatography [55]. He detected a variety of amino acids, some of which were proteinogenic and some of which

were not; subsequent analyses of his reaction samples with more sophisticated instrumentation have revealed additional amino acid products (Figure 1.6) [55, 130].

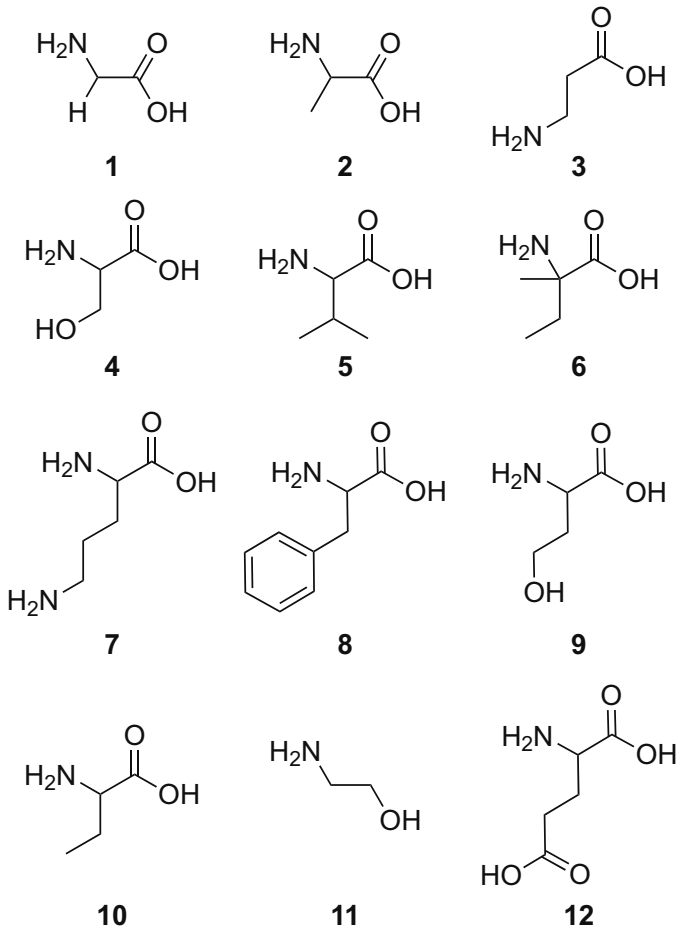

Figure 1.6 Selected Miller-Urey reaction products. Amino acids and other chiral compounds were formed in racemic mixtures. **1** glycine, **2** alanine, **3** β-alanine, **4** serine, **5** valine, **6** isovaline, **7** ornithine, **8** phenylalanine, **9** homoserine, **10** α-aminobutyric acid, **11** ethanolamine, **12** glutamic acid.

1.4.6.2 The Formose Reaction

The formose reaction occurs when formaldehyde is treated with a base [48] and was first described by Butlerov in 1861 [131]. This reaction is undoubtedly the best-known route to prebiotic sugar synthesis and proceeds via a series of aldol condensations

and tautomerizations to yield first glycolaldehyde, then trioses, tetroses, and larger molecules (Figure 1.7) [48].

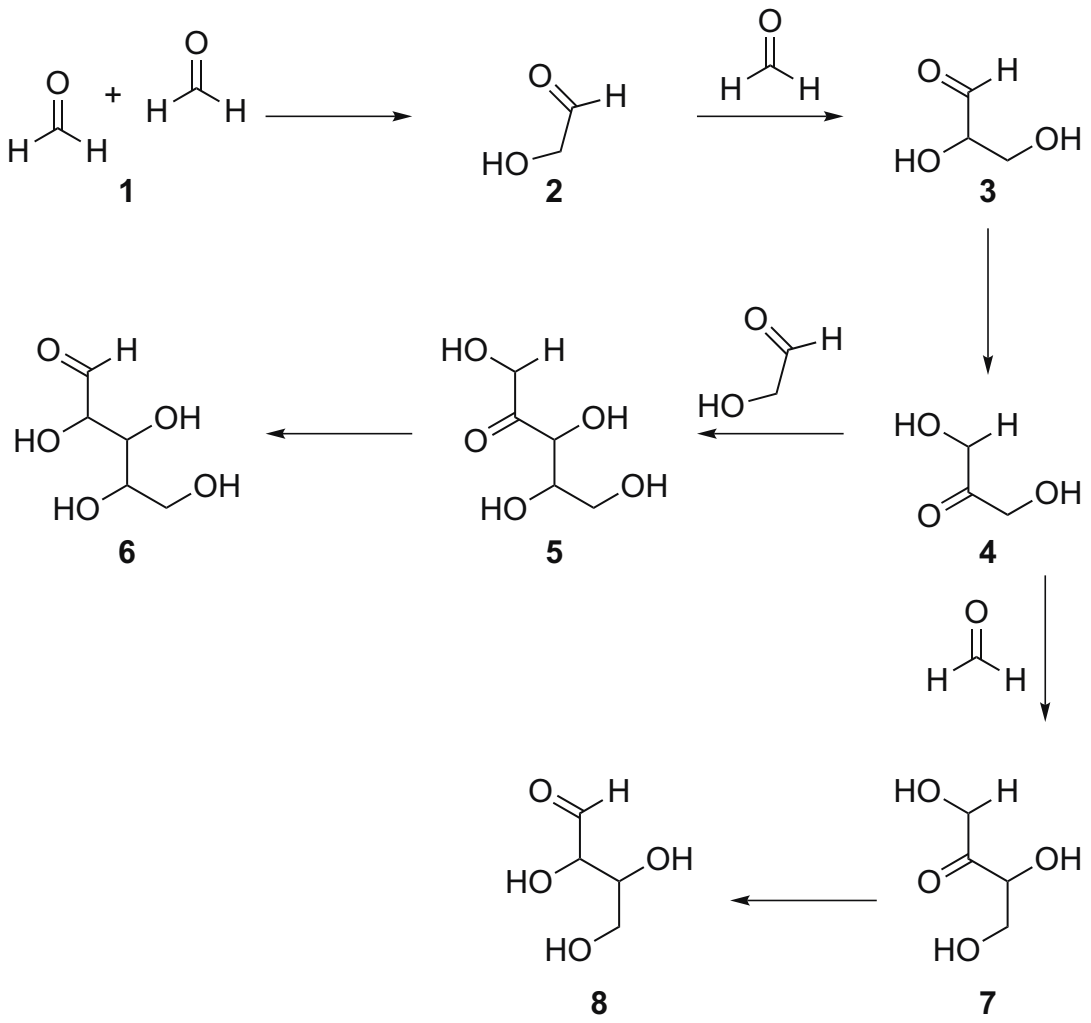

Figure 1.7 The formose reaction. **1** formaldehyde, **2** glycolaldehyde, **3** glyceraldehyde, **4** dihydroxyacetone, **5** ketopentoses, **6** aldopentoses, **7** erythrulose, **8** aldotetroses. The formose reaction is base-catalyzed and both L and D sugars are formed. Structures shown may react further to generate additional products.

The formose reaction is autocatalytic with a slow first step (condensation of two formaldehyde molecules); the presence of increased levels of glycolaldehyde leads to a corresponding spike in reaction rates [48]. In the 150 years since Butlerov initially

published his results, this reaction has generated continuing interest, with investigations of amine and amino acid catalysis [47], phosphate mineral catalysis [46], and ultraviolet light coupled with zeolite catalysis [132]. Yields of ribose generated in the formose reaction remain low and the reaction mixture remains complex [133]; however, reactions conducted in the presence of the Pb^{2+} ion have led to aldopentose yields as high as 30% [134]. The possibility remains that ribose was not the sugar unit of the earliest nucleic acid analogs, but gradually came to replace other monosaccharides as more efficient syntheses were made possible in the increasingly organized prebiotic chemical environment.

1.4.6.3 Prebiotic Syntheses of Pyrimidines

While several facile prebiotic syntheses of purines have been reported [40, 107], prebiotic routes to pyrimidines have proven to be more difficult. These syntheses frequently require high-energy reactants such as cyanoacetaldehyde, a compound which has been detected in spark-discharge experiments (Figure 1.8) [102, 135, 136]. A gentler photochemical synthesis has been reported for uracil and thymine in the presence of montmorillonites (Figure 1.8) [137].

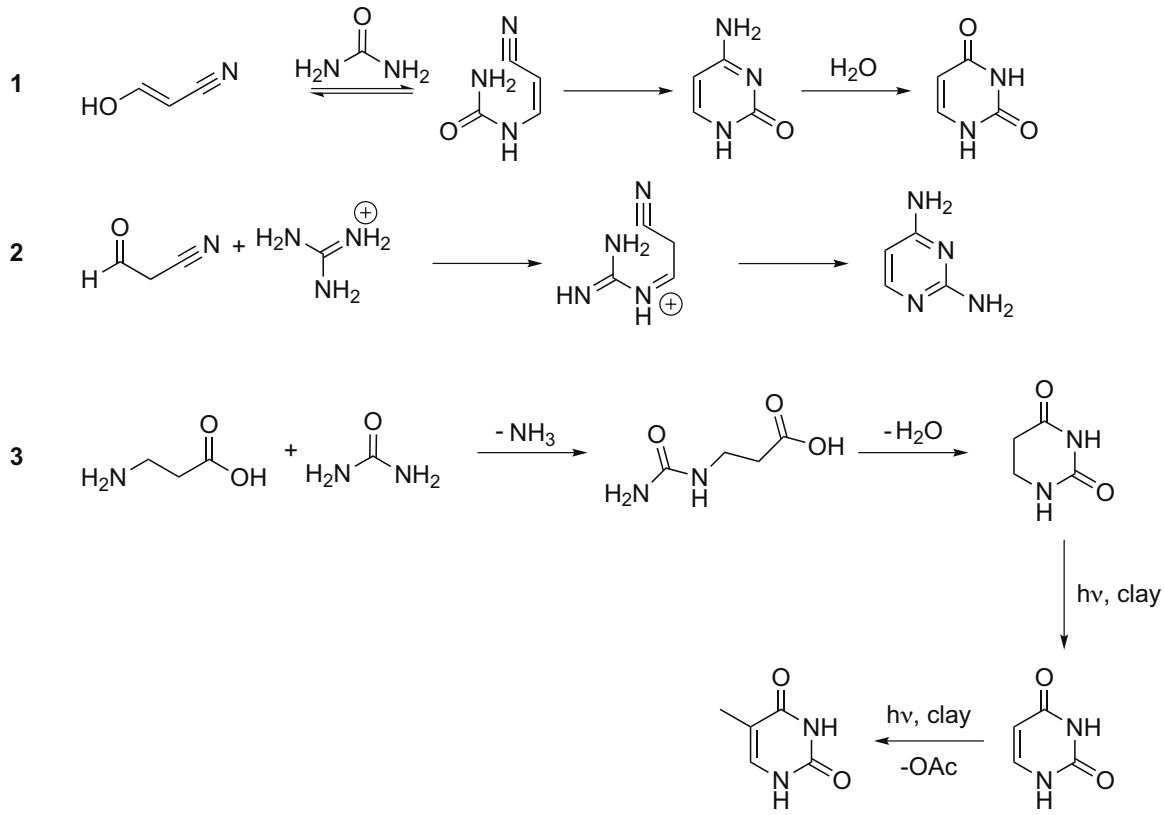

Figure 1.8 Prebiotic syntheses of pyrimidines. In reaction **1** [136], cyanoacetaldehyde (shown here as a tautomer) reacts with urea to form cytosine, which can be hydrolyzed to uracil. In reaction **2** [135], cyanoacetaldehyde reacts with guanidine to generate small amounts of 2,4-diaminopyrimidine. In reaction **3** [137], β-alanine and urea react to form 5,6-dihydrouracil which can be treated with ultraviolet light in the presence of clays to generate uracil and thymine.

1.4.7 Exogenous Delivery of Compounds from Meteorites and Interplanetary Dust Particles

We have presumed that the reactions described thus far occurred on the early Earth in the presence of simple precursors such as HCN, urea, and formamide, or of their hydrolysis or reaction products including ammonium formate, formaldehyde, and cyanoacetaldehyde. It is also possible, however, that related processes could have occurred on extraterrestrial bodies, with products being delivered to Earth on meteorites.

Many biologically-relevant products have been detected in meteorites (Figure 1.9); these include purine nucleobases [119]; sugars and sugar acids [138]; and amino acids [139]. Chiral compounds among these are detected as racemic mixtures, and many molecules are found which are not employed by contemporary life forms; this suggests the synthesis of these products on meteorite parent bodies. Stable isotope analysis can assist in establishing an extraterrestrial origin for achiral molecules [119, 138, 139].

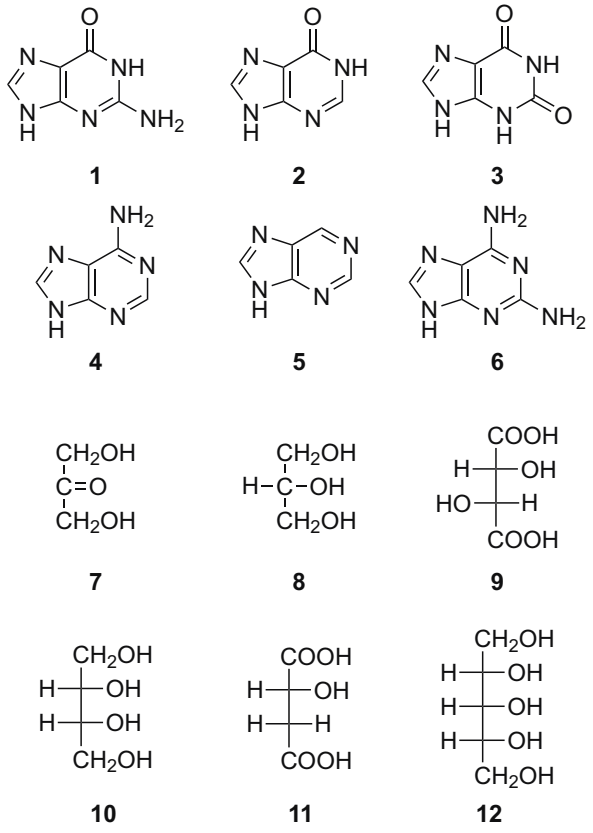

Figure 1.9 Selected nucleobases, sugars, sugar alcohols, and sugar acids detected in meteorites. Amino acids are also found in meteorites; many of these are the same as those reported in the Miller-Urey experiment (see Figure 1.6). Amino acids and sugar compounds in meteorites are either racemic mixtures of L and D compounds, or display slight enantiomeric excesses of one or the other chiralities. **1** guanine, **2** hypoxanthine, **3** xanthine, **4** adenine, **5** purine, **6** 2,6-diaminopurine, **7** dihydroxyacetone, **8** glycerol, **9** tartaric and mesotartaric acid, **10** erythritol and threitol, **11** malic acid, **12** ribitol and isomers [119, 138].

It should be noted, however, that several orders of magnitude less organic material is estimated to have been delivered by meteorites than by interplanetary dust particles (IDPs) [140]. Further, the inventories of compounds available from each of these sources would have been somewhat different, with IDPs being rich in polycyclic aromatic hydrocarbons and meteorites containing nitrogen heterocycles and more oxidized forms of carbon [140]. However, some syntheses of biologically relevant molecules have been reported from compounds similar to those believed to have been present on IDPs, and moreover, the importance of various endogenous and exogenous sources of biomolecules may be highly dependent on early Earth's atmospheric composition [140].

1.5 ASSEMBLING COMPONENTS FROM THE PREBIOTIC CHEMICAL INVENTORY INTO THE FIRST NUCLEIC ACIDS

While many plausibly prebiotic syntheses of the individual components of nucleic acids have been reported, and while delivery of these same compounds by meteorites of IDPs is also possible, it is not trivial to assemble these components into functional biopolymers. Many scientists have addressed aspects of this problem, taking a variety of approaches and starting from a variety of initial assumptions.

1.5.1 Nucleosidation and Phosphorylation

Two main strategies exist for forming a nucleotide under plausibly prebiotic conditions. The first requires preexisting supplies of sugars and nucleobases, and later a phosphate source. In the second, the entire nucleotide unit is built up from simple precursors as opposed to the individual synthesis of components.

1.5.1.1 Dehydration/Condensation Reactions

Prebiotic syntheses of nucleobases and sugars have been demonstrated. The key challenge to adding nucleobases to sugars is that this reaction is a dehydration (condensation) reaction: water is produced. According to Le Chatelier's principle, in an aqueous environment, the equilibrium for a reaction that generates water will be shifted in the direction of water uptake, i.e. hydrolysis, rather than water release. The primary strategy that has been invoked to address this problem is the drying reaction. Fuller and colleagues described successful syntheses of inosine and adenosine from solutions or ribose, salts, and hypoxanthine or adenine, respectively (Figure 1.10) [141]. Guanosine synthesis was likely not achieved (or detected) due to guanine's low solubility in water, and similar reactions involving the canonical pyrimidine nucleobases have also been unsuccessful [141]. Formation of the nucleoside zebularine, containing the noncanonical base 2-pyrimidinone, has been reported in a dehydration/condensation reaction (Figure 1.10) [142].

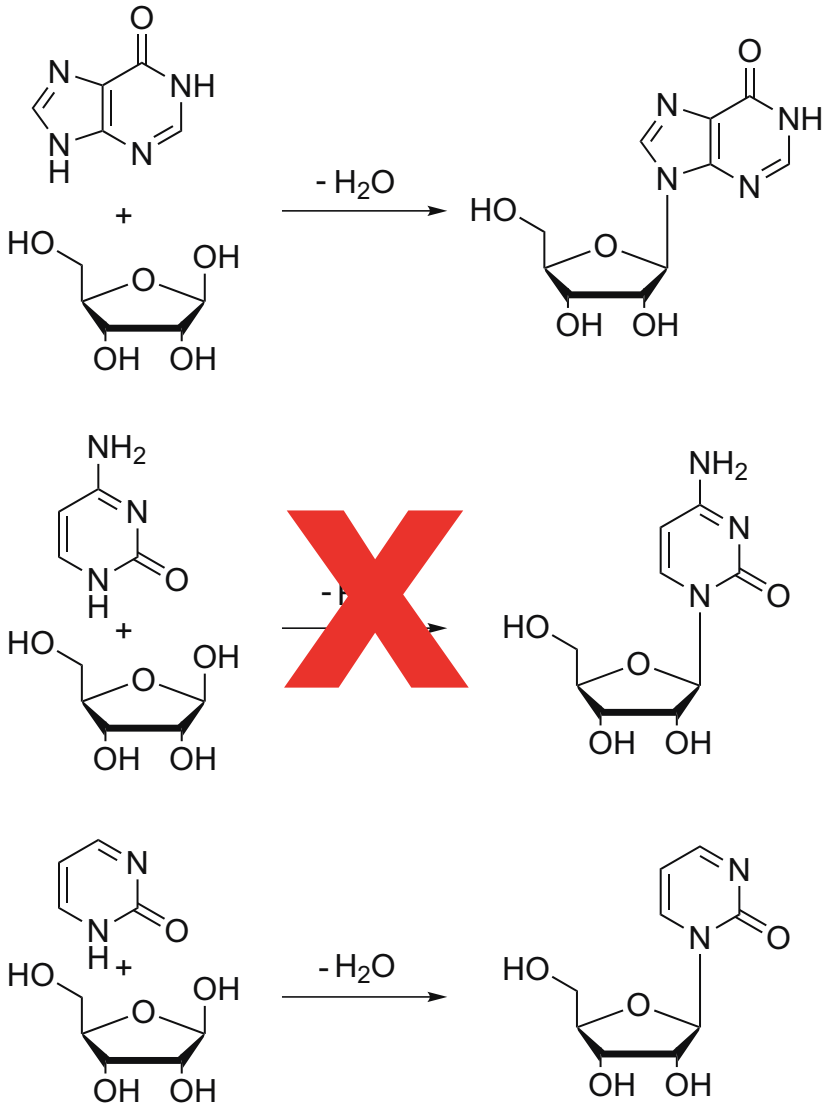

Figure 1.10 Nucleoside formation in a condensation/dehydration reaction. Top row: purines such as hypoxanthine (pictured) and adenine form nucleosides in drying reactions. Middle row: canonical pyrimidines such as cytosine (pictured) and uracil do not add to ribose. Bottom row: some noncanonical pyrimidines, e.g. 2-pyrimidinone, have been shown to form nucleosides in drying reactions.

1.5.1.2 Phosphorylation of Nucleosides

Once a nucleoside has been successfully formed, it must be linked together with other nucleosides into a chain; DNA and RNA employ phosphate groups for this purpose. Various attempts to phosphorylate nucleosides have been undertaken. Since this is also a

dehydration/condensation reaction, one popular approach has been to perform phosphorylations in non-aqueous solvents, including formamide. Addition of phosphate to a nucleoside can be accomplished by mild heating in formamide in the presence of a phosphate source, which could be a simple potassium phosphate [83, 84] or a more complex phosphate mineral [81]. The resulting mixtures of nucleotides may contain species with phosphate moieties attached at the biologically irrelevant 2′ position of the sugar (Figure 1.11) [83, 84]. Phosphorylations aided by cyanamide have also been reported [143].

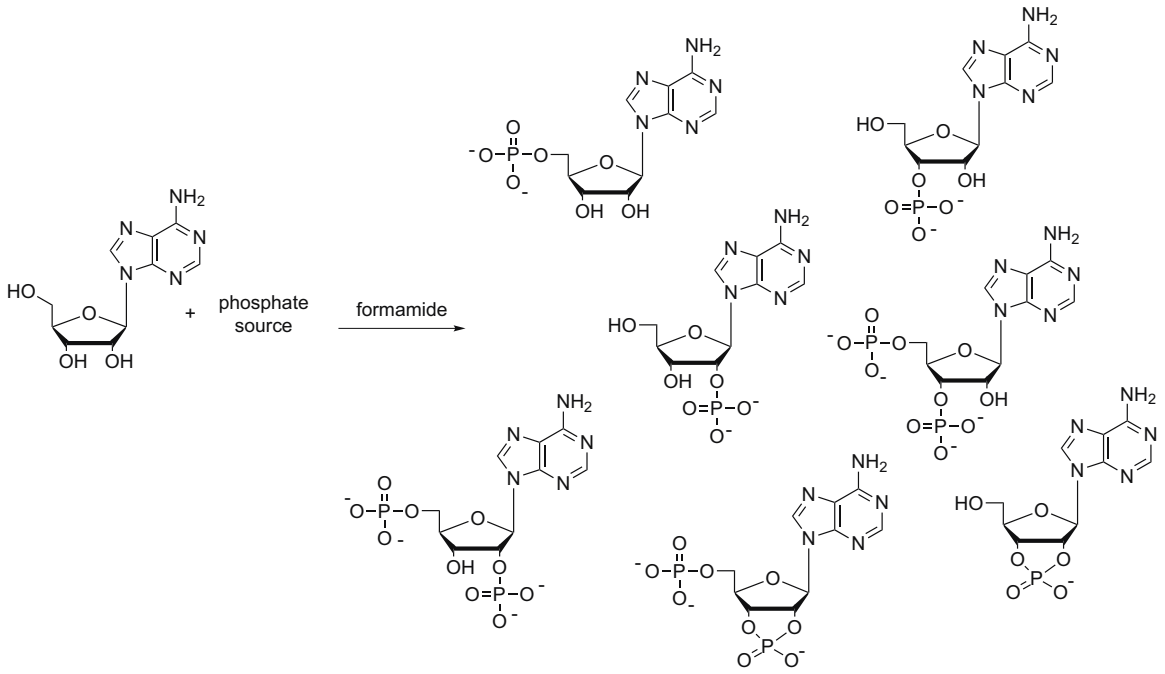

Figure 1.11 Nucleosides can be phosphorylated in formamide solutions in the presence of phosphate sources. Phosphate sources include, but are not limited to, inorganic condensed phosphates and polyphosphates as well as phosphate minerals. Various mono-, di-, and cyclic phosphates are generated in these reactions [83].

1.5.1.3 Synthesis of Activated Nucleotides

An alternative approach to nucleotide synthesis was demonstrated by John Sutherland and coworkers; this strategy uses simple precursors to build a sugar scaffold upon which the nucleobase is constructed in a stepwise fashion (Figure 1.12) [144].

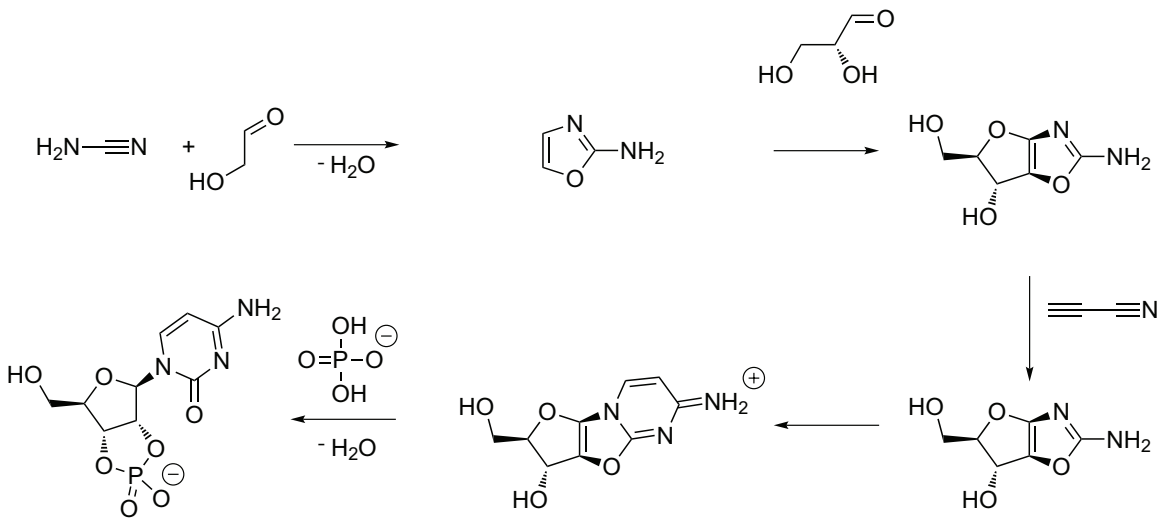

Figure 1.12 Prebiotically plausible synthesis of activated pyrimidine nucleotide [144].

This approach has several advantages: this chemistry can be carried out in water, D-ribose with the correct furanose ring form is generated, and the process results in a nucleotide containing cytosine, one of the canonical pyrimidine bases [144]. However, this directed approach may not be evolutionarily sound. Whereas nucleobases, particularly purines, are known to assemble by stacking in aqueous solutions [145], and nucleic acid double helices are held together in large part because of hydrogen-bond interactions between purine and pyrimidine bases, few of the intermediates in Sutherland's synthesis possess aromatic stacking surfaces and most also appear to be incapable of hydrogen bonding or orienting themselves to form stable base pairs [144].

Thus, the likelihood of assembly of Sutherland's intermediates into more complex structures is low and there is a lack of selective pressure to preserve these compounds in a complex prebiotic environment. It is also improbable that the same nucleotides we make use of today were the only ones that formed in a primordial soup; a more likely scenario is a nucleic-acid-like polymer that contained a variety of easy-to-synthesize and easy-to-assemble elements that looked quite different from what we recognize today, but which evolved into RNA over time.

1.5.2 Polymerization

Regardless of how nucleosides and nucleotides are formed, these molecules were connected to one another to construct functional biopolymers. Several strategies exist for accomplishing this; these include polymerization of activated nucleotides and the use of easier-to-synthesize alternative backbones.

1.5.2.1 Polymerization of Activated Nucleotides

One tactic for stitching nucleotides together involves activating 5′ nucleoside monophosphates (Figure 1.13); these are able to polymerize in solution or on a montmorillonite clay surface [86]. Reactions on clay surfaces result in oligomers up to 55 bases long, while solution-phase reactions result in much shorter polymers [86].

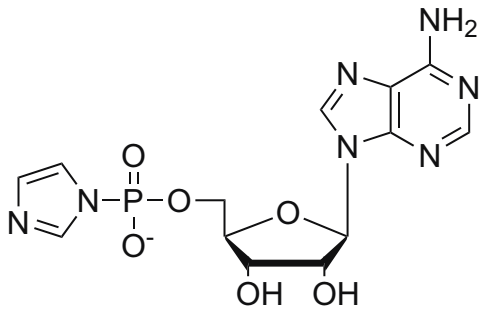

Figure 1.13 Imidazole-activated RNA mononucleotide.

While clay surfaces are attractive because they assist in the generation of longer nucleic acid polymers, adhesion of oligonucleotides to a clay surface may actually disfavor replication – a double helix would not be able to form with one strand's backbone held in a linear conformation by the montmorillonite. The development of a more robust solution-phase polymerization would avoid this problem (see Section 1.5.2.3).

1.5.2.2 Alternative Backbones

Numerous alternative backbones have been proposed for primordial genetic polymers; some of these make use of phosphate linkages, though others do not (Figure 1.14).

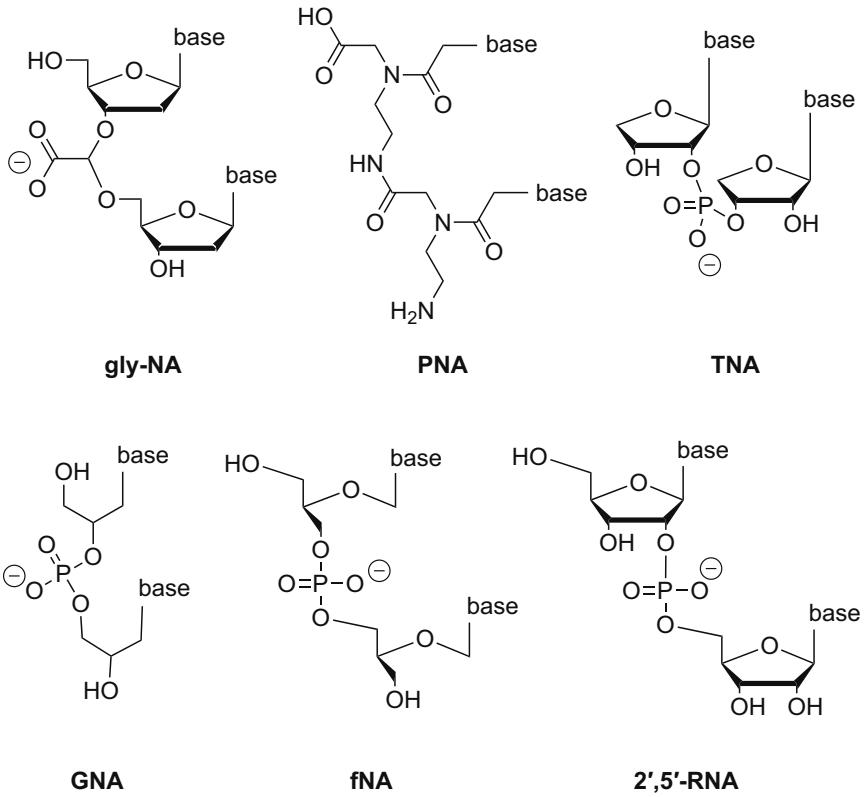

Figure 1.14 A selection of alternative backbones proposed as possibilities for pre-RNA nucleic acid analogs.

Among the phosphate-linked alternatives, various pairing and replicative abilities have been demonstrated. Threose nucleic acid (TNA) is able to base pair with natural RNA and DNA and has been shown to template DNA synthesis in the presence of a promiscuous DNA polymerase [146, 147]. Glycol nucleic acid (GNA) possesses the same backbone length and base orientation as TNA, but lacks TNA's sugar ring. GNA monomers have (*R*) and (*S*) forms which are able to pair with other GNA strands of their own chirality and a complementary base sequence; additionally, (*S*)-GNA can cross-pair with RNA [148]. Flexible nucleic acid (fNA) uses a glycerol derivative for its sugar and mimics the connectivity of RNA or DNA while lacking a 2′ carbon; however, the inclusion of multiple fNA monomers in an oligonucleotide chain results in a greatly destabilized helix [149]. 2′,5′-RNA can form duplexes with itself and with natural RNA, but not with DNA except at very high salt concentrations [150].

Non-phosphate-linked nucleic acid analogs vary widely in their structures and properties. Peptide nucleic acid (PNA), while it replaces the sugar-phosphate backbone entirely, forms stable heteroduplexes with DNA, as well as extremely stable duplexes with itself [151]. Glyoxylate-linked nucleic acid (gly-NA) dinucleotides have been synthesized in model prebiotic drying reactions; the glyoxylate linkage is attractive because glyoxylate has similar size and charge as phosphate, because free glyoxylate is likely to have been available on the early Earth, and because the formation of this linkage does not require activation of nucleosides [152].

The characteristics of pre-RNA nucleic acids have been the subject of much speculation and many experiments have been performed to assess the fitness of the various backbones that have been proposed. Since it is not possible to know the chemical

identity of the first functional biopolymer, scientists must instead approach the topic with an eye towards determining which of the alternatives might have been the easiest to assemble by abiotic reactions and which could also base pair to form a duplex, thereby having the potential to replicate in a semi-conservative manner.

1.5.2.3 Template-Directed Ligation

A solution of nucleotides with activated 5′ ends will generate some short polymers, but an upper limit of a ten-nucleobase oligomer is soon reached [86]. When an activated oligonucleotide chain becomes long enough, cyclization can occur (Figure 1.15, Panel A). Once an oligomer has cyclized, further chain elongation becomes impossible. Additionally, this approach results in the generation of a modest number of polymers with a large percentage of unreacted or hydrolyzed starting material remaining in solution at the time of analysis [86, 153].

One way to circumvent the strand cyclization problem is with the use of intercalating small molecules [154]. Intercalation unwinds and stiffens a nucleic acid double helix, preventing the 3′ and 5′ ends of the same strand from coming into contact with one another and thus inhibiting cyclization. If one begins not with monomers but instead with a pool of tetramers possessing two-base-pair complementarity and two-base overhang, it is possible to construct a system whereby these short oligonucleotides tile with one another via their overhanging bases (Figure 1.15, Panel B). In the presence of an activating chemistry and an intercalator, polymers of up to one hundred bases can be generated [154]; this nearly doubles the lengths achieved by activated monomers on montmorillonite surfaces [86] and results in the conversion of most starting material to product [154].

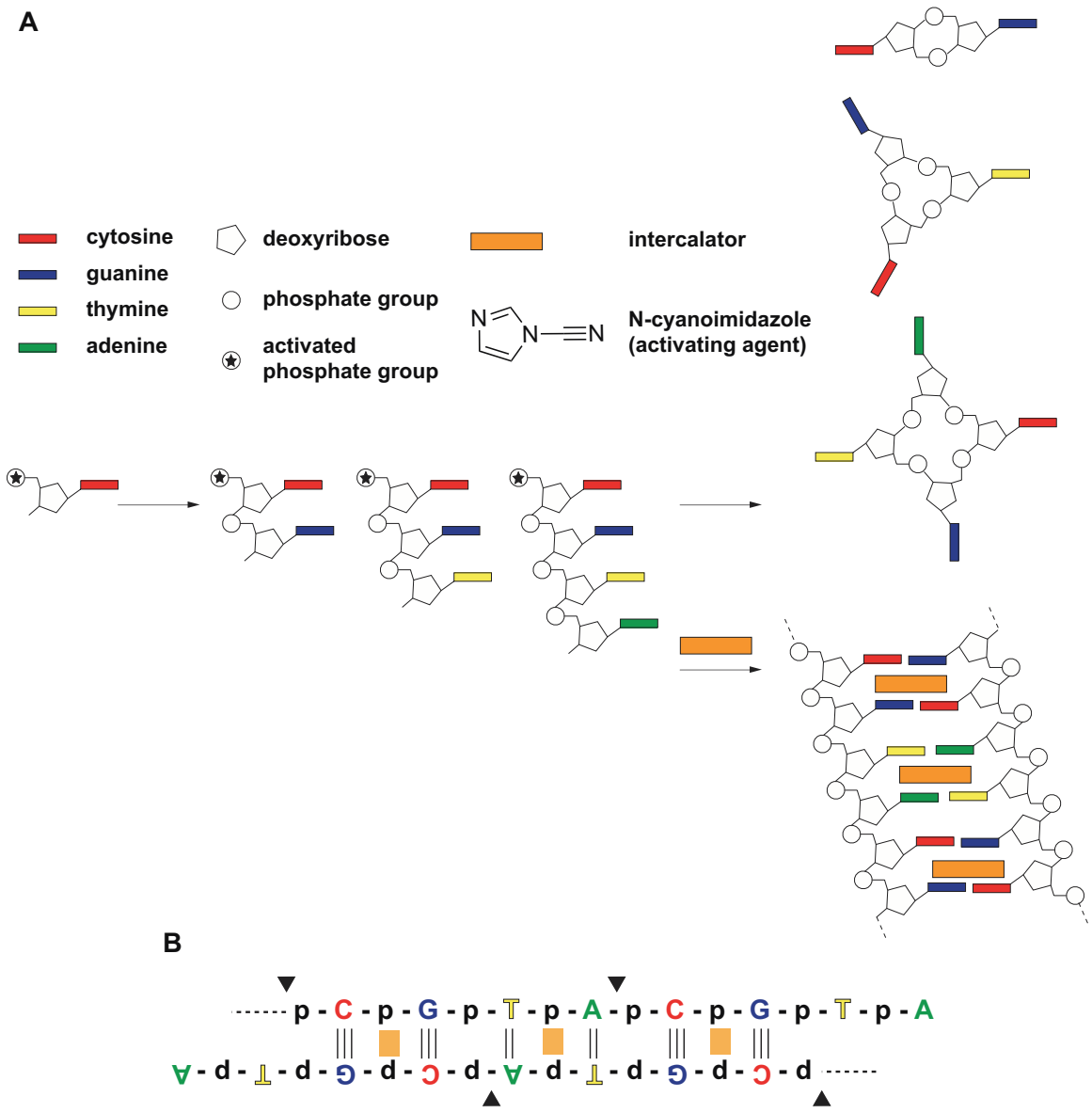

Figure 1.15 (A) Activated nucleotides polymerize in aqueous solution, but the resulting oligonucleotides tend to cyclize at a certain length, which may be as short as two, three, or four bases. One way to prevent this cyclization is to add an intercalator, which will stack between base pairs, preventing the ends of the polymer chain from contacting one another. (B) In the presence of an intercalator, 5′-phosphorylated DNA tetramers can assemble into long duplexes in aqueous solution if the sequence of these tetramers is such that it contains two base pairs and a two-base-pair overhang. Black triangles indicate nicks in the duplex backbone, and vertical lines represent hydrogen bonds between the bases in a pair. Addition of N-cyanoimidazole activates the 5′ phosphates for polymerization with the 3′ hydroxyls of neighboring tetramers. Figure modeled after reference [154].

1.6 NUCLEIC ACIDS WITH PURINE-PURINE BASE PAIRS

The first nucleic acids likely possessed different sugars and backbone linkages than contemporary DNA or RNA. It is equally possible, however, that they possessed alternative recognition elements. One idea that has been proposed is that the first nucleic acids contained only purine bases [24, 45]. Purine nucleobases stack more favorably in aqueous solution than are pyrimidines [145], and a double helix of polyadenylic acid and polyinosinic acid was proposed by Alex Rich as early as 1958 [155].

1.6.1 Purine-Purine Base-Pairing Structures

A variety of purine-purine base-pairing motifs have been demonstrated in the literature. These can include duplexes, triplexes, quadruplexes, and even pentaplexes. While some of these polypurine structures require specific solution conditions or the presence of additional small molecules, others form spontaneously.

1.6.1.1 Poly(A) Self-Structure

Polyadenosine self-structures have received much attention because of the poly(A) tails that have been demonstrated to exist in messenger RNAs (mRNA) in eukaryotic cells. At least two polyadenylic acid structures have been reported (Figure 1.16); one exists in solution at acidic pH [156] and the other occurs at neutral pH and is induced by the presence of selected small molecules [157].

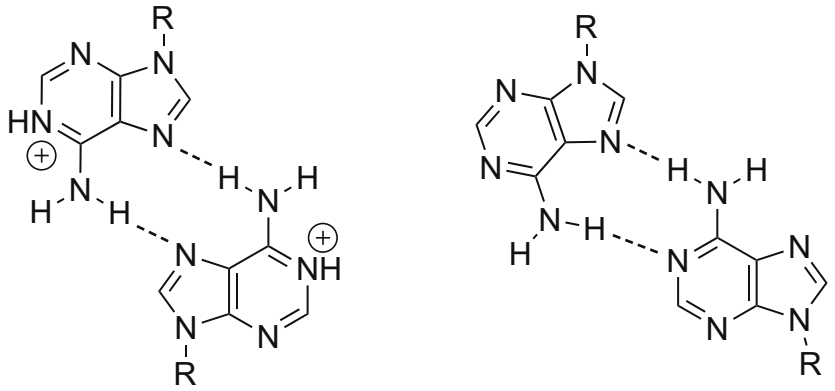

Figure 1.16 Poly(A) self structures. Left: A·A structure formed at acidic pH [156]. Right: A·A induced at neutral pH by the presence of certain small molecules [157, 158]. R represents a sugar unit.

1.6.1.2 G-Quadruplexes

The G-quartet (Figure 1.17) is another polypurine structure with biological relevance. G-quadruplexes have not been demonstrated *in vivo*, but quadruplex-forming sequences are found in telomeres and the promoter regions of genes [159].

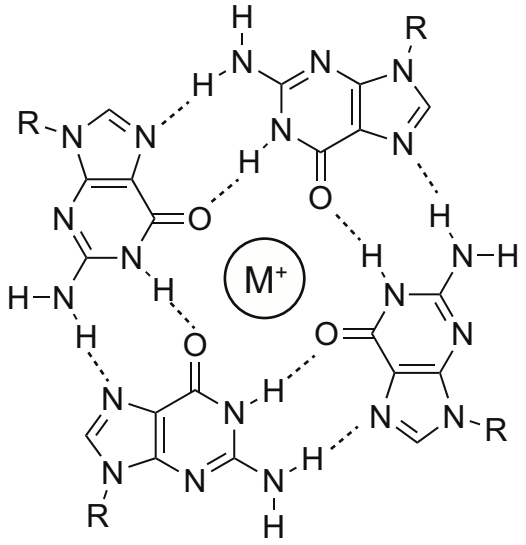

Figure 1.17 G-quartet. A cation is required for the formation of this structure and may be ammonium or a Group I metal. Lithium cannot be easily dehydrated and thus does not support quadruplex formation. R represents a sugar unit.

G-quartets consist of four guanine bases arranged around a central ion channel; most alkali metals as well as the ammonium ion can promote the formation of these structures, which exist in various forms depending on solution conditions as well as nucleotide sequence [159].

1.6.1.3 Other Polypurine Structures

Additional polypurine structures have been reported or proposed based on biophysical characterizations (Figure 1.18). These include a triplex with two polyinosinic acid strands and one polyadenylic acid strand (Figure 1.18, Panel A) [160], a triplex of polyxanthylic acid and polyadenylic acid (Figure 1.18, Panel B) [161], a pentaplex composed of isoguanine residues (Figure 1.18, Panel C) [162], and a polyxanthylic acid duplex in which the bases may exist in one of several tautomeric forms (Figure 1.18, Panel D) [163].

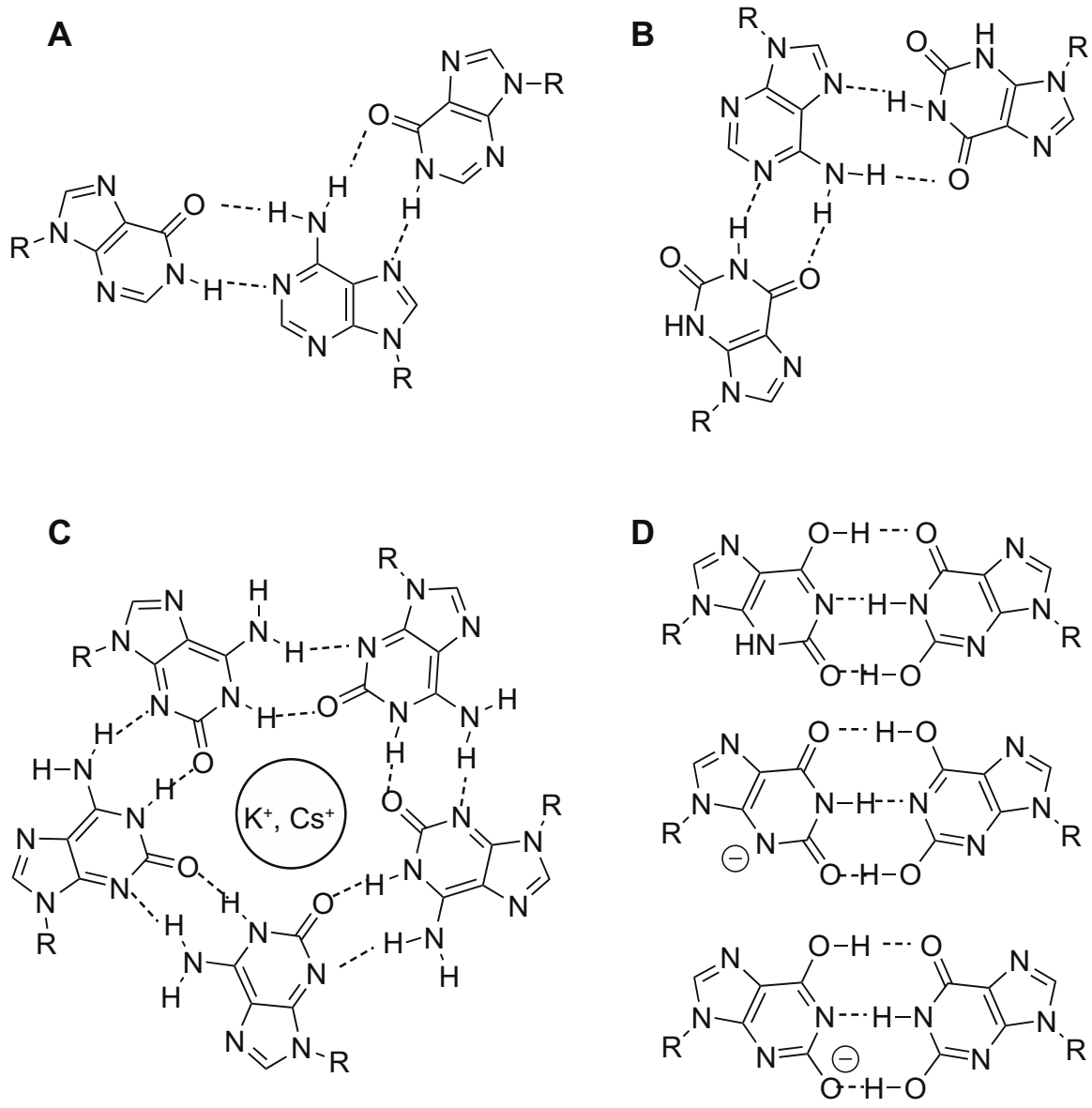

Figure 1.18 Homopurine base pairing motifs. (A) I·A·I triplet from a crystal structure determined by Arnott and Bond [160]. (B) X·A·X triplet proposed by Fikus and Shugar [161]. (C) Isoguanine quintet structure proposed by Chaput and Switzer [162]. (D) Three possible hydrogen bonding patterns for the X·X base pair from the crystal structure by Arnott *et al* [163]. R represents a sugar unit.

1.6.2 Purine-Purine Pairs Inserted into Watson-Crick DNA Structures

Purine-purine base pairs (Figure 1.19) have also been inserted into otherwise normal Watson-Crick DNA structures. Examples of these include the adenosine·inosine pair in the self-complementary DNA duplex 5′-GGIACC-3′ [164] and the 2,6-diaminopurine·xanthosine pair in homo-DNA, a hexapyranosyl 6′,4′ linked nucleic acid analogue developed by the Eschenmoser group [165].

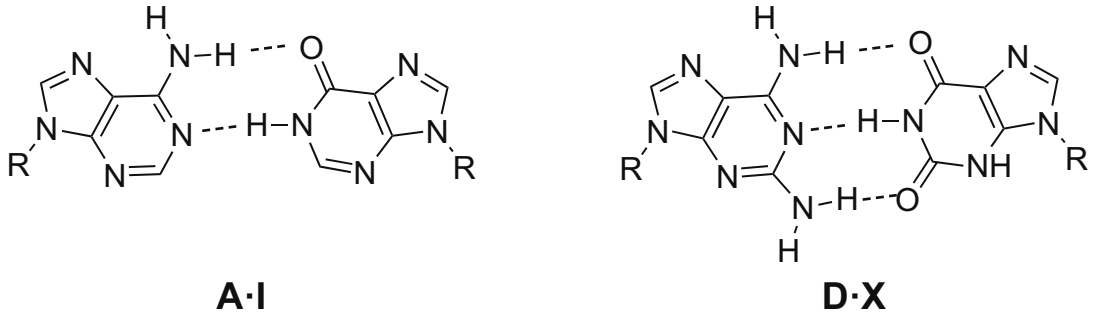

Figure 1.19 The adenine·inosine (left) and diaminopurine·xanthine (right) base pairs. R represents a sugar unit.

1.6.3 Antiparallel DNA Duplexes Containing Only Purine Bases

Additionally, antiparallel DNA duplexes containing all purine bases have recently been reported. Battersby and co-workers have demonstrated duplex formation in short, mixed-sequence oligonucleotides containing adenosine·inosine and guanosine·isoguanosine base pairs (Figure 1.20) [44]; Heuberger and Switzer substituted the 2,6-diaminopurine·7-deazaxanthosine base pair (Figure 1.20) for adenosine·inosine with the object of increasing stability via the inclusion of additional hydrogen bonds between nucleobase recognition elements [43].

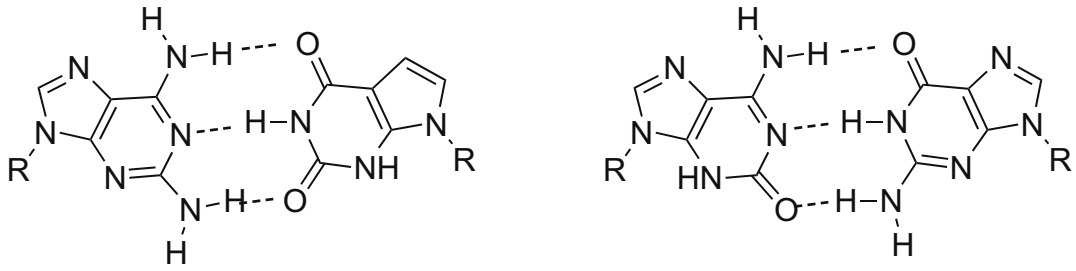

Figure 1.20 The diaminopurine·7-deazaxanthine (left) and isoguanine·guanine (right) base pairs. R represents a sugar unit.

1.7 SMALL MOLECULE INTERACTION WITH NUCLEIC ACIDS

Small molecule interactions with nucleic acids have generated a great deal of interest in the medical field. Some cancer drugs such as the nitrogen mustard melphalan [166], the nitrosourea carmustine [167], and the alkyl sulfonate busulfan [168] alter DNA by alkylation of the nitrogenous bases; others, such cisplatin or calicheamicin, bind irreversibly to DNA or cause backbone cleavage, respectively [169, 170]. Small molecules which noncovalently bind to DNA are also used as medical treatments; these include doxorubicin for a variety of cancers [171], berenil against trypanosomatid protozoans [172], and netropsin for certain viruses [173]. Many DNA-binding small molecules also have mutagenic properties; these include etoposide, camptothecin, and mitomycin [174, 175] as well as many of the previously mentioned antineoplastic drugs.

1.7.1 Noncovalent Binding Modes

Several modes of noncovalent binding of small molecules to DNA have been characterized; these include intercalation and groove binding [176]. Hydrogen bonding, base stacking, van der Waals interactions, and release of water and/or cations are

important driving forces in these binding interactions which do not alter the chemical structure of the DNA bases or backbone.

1.7.1.1 Intercalation

Molecules that intercalate nucleic acids are planar, often aromatic, and frequently also positively charged; their planar portions are inserted between the base pairs (Figure 1.21), causing the double helix to unwind 10° to 20° [177], depending on the identity of the small molecule. Some drugs, such as ethidium bromide, cause an even greater degree of unwinding [178]. This unwinding is accompanied by lengthening of the helix as base pairs unstack to accommodate the intercalator in the binding site [177]. Intercalation is frequently accompanied by changes in sugar conformation or backbone torsion angle and may distort the conformation of nucleotides several bases away from the intercalation site [177]. Some intercalating molecules exhibit specific nucleic acid sequence preferences (e.g. actinomycin D binds DNA only at GpC steps) [177].

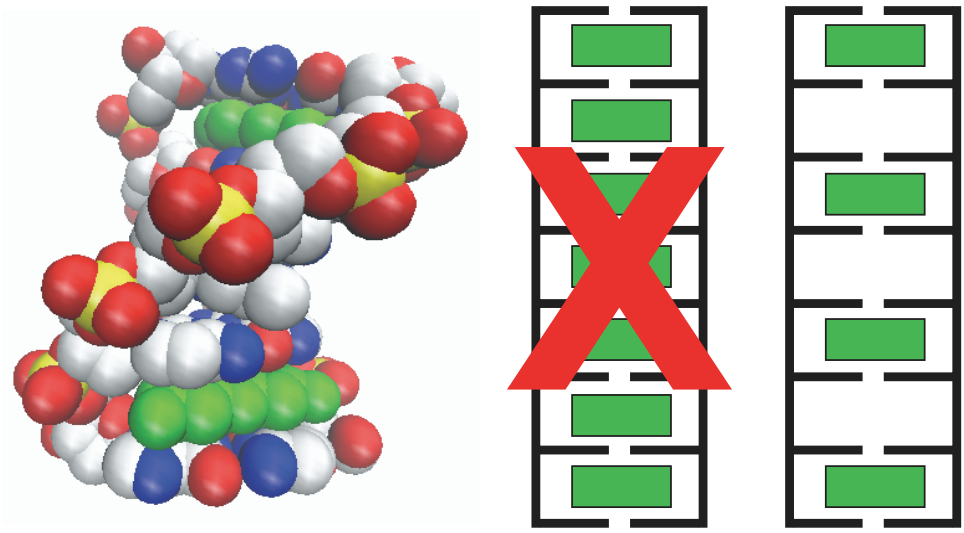

Figure 1.21 Intercalation. Left: Intercalator proflavine complexed with self-complementary DNA duplex 5′-CGATCG-3′. From PDB structure 3FT6 [179]. DNA carbon atoms are white, oxygen atoms red, nitrogen atoms blue, and phosphorus atoms yellow. Proflavine molecule is green. Intercalation unwinds and lengthens the DNA double helix. Center: Violation of the nearest neighbor exclusion principle. Right: Nearest neighbor exclusion principle, which states that only one intercalator molecule can bind for every two base pairs.

The nearest neighbor exclusion principle holds that the maximum loading of an intercalator into a nucleic acid duplex is one small molecule for every two base pairs [177]; that is, that only every other potential binding site can be occupied (Figure 1.21). Various explanations have been suggested for this phenomenon: steric effects due to the structural changes that accompany intercalation [180], a reduction of negative electrostatic potential near the binding site rendering further electrostatic interactions unlikely [180], and vibrational entropy favoring the more flexible semi-intercalated double helix over the more rigid fully-intercalated structure [181]. The actual physical basis for the nearest neighbor exclusion principle, however, remains unresolved.

1.7.1.2 Minor Groove Binding

Some molecules interact with the minor groove of DNA; these include netropsin, distamycin A, and Hoechst 33258 (Figure 1.22) [182]. These drugs exclusively bind to the narrow groove of (A·T) tracts in DNA [173, 183]. Minor groove binding is accompanied by changes in solvent interactions, as well as by cation release if the groove binder is positively charged; there is thus a strong entropic contribution to this type of DNA-small molecule interaction [176].

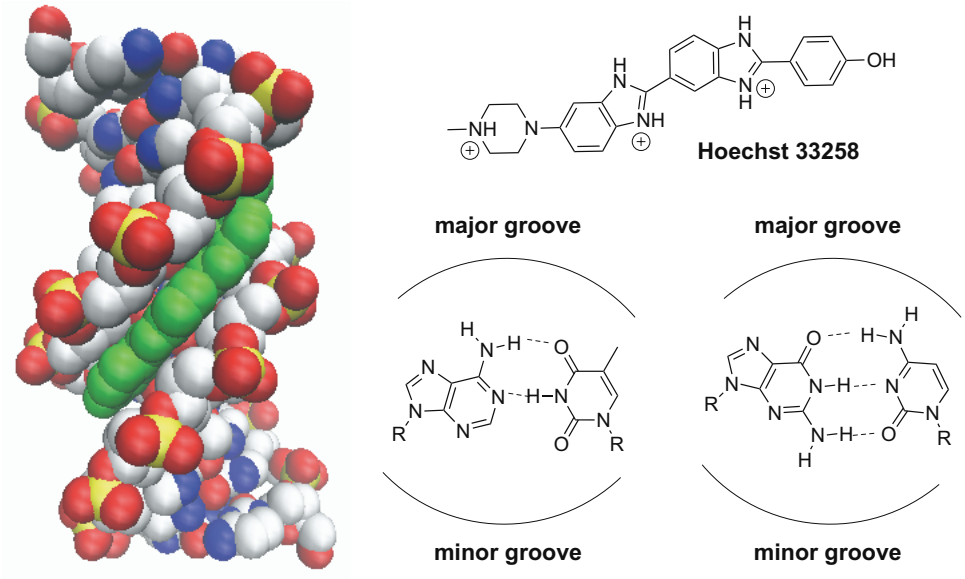

Figure 1.22 Minor groove binding. Left: Groove-binding dye Hoechst 33258 complexed with self-complementary DNA duplex 5′-CGCAAATTTGCG-3′. From PDB structure 264D [184]. DNA carbon atoms are white, oxygen atoms red, nitrogen atoms blue, and phosphorus atoms yellow. Hoechst 33258 molecule is green. Right (top): Chemical structure of Hoechst 33258. Right (bottom): Chemical structures of the A·T (bottom center) and G·C (bottom right) base pairs of DNA, with major and minor groove positions indicated. R represents a sugar unit.

Additionally, some molecules which are known to intercalate also possess moieties that project into and interact with the minor groove of DNA; these compounds include actinomycin D [185], daunomycin [186], and ethidium [187].

1.7.2 The Molecular Midwife Hypothesis

Hud and Anet first proposed the molecular midwife hypothesis in 2000 as a way to explain the assembly of the first nucleic acids from prebiotically-synthesized nucleobase units [188]. While purine bases, in particular, are known to form stacks in aqueous solution [145], some base-pairing scheme is needed for information transfer. Hud and Anet's proposal posits the existence of a planar molecule with a size and shape similar to a base pair – or tetrad – which can stack with that base pair. Repeating stacks of such units (Figure 1.23) would facilitate hydrogen bonding between paired bases due to the creation of a localized water-free environment, would allow glycosylation of bases at the edges of stacks, and would pre-position nucleosides in an arrangement favorable to eventual backbone formation [188].

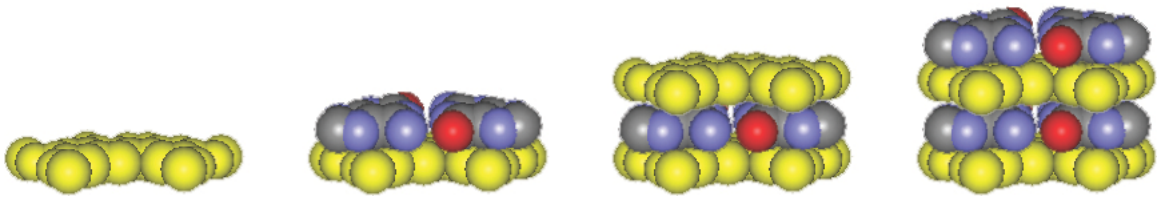

Figure 1.23 Assemblies illustrating the molecular midwife hypothesis. A planar molecule, represented here in yellow, provides a stacking surface for a guanine·isoguanine base pair (carbon represented as gray, oxygen as red, nitrogen as purple). Another midwife molecule stacks on top of the base pair, and the process continues until longer alternating midwife-base pair stacks are present in solution. Different purine·purine base pairs can assemble in the same stack, and stacks with smaller (e.g. Watson-Crick base pairs) or larger (e.g. G-quadruplexes) may also form. These stacking interactions assemble base pairs in the proper orientation for backbone formation. Intercalation may also facilitate information transfer through semi-conservative replication of base sequences [188].

A midwife molecule could easily be imagined to be synthesized from prebiotically plausible components via formose-type or HCN reactions, or some combination of the two processes [188].

1.7.3 From Intercalation to Putative Prebiotic Midwife Molecules

While many intercalators have been isolated from plant cells or microbes, these molecules are the end products of complex biosynthetic pathways which would not have been accessible on the early Earth. However, present-day intercalators are able to serve as proxies for putative molecular midwives; therefore, the study of intercalator interactions with potentially primitive nucleic acids can provide us with valuable clues as to the chemical natures and structural frameworks of early midwife molecules.

1.8 BRIDGING THE GAP BETWEEN SYNTHESIS OF SMALL MOLECULES AND FORMATION OF LARGER ASSEMBLIES

Prebiotic syntheses of nucleobases and sugars are well characterized. Some degree of success has been achieved at forming nucleosides and nucleotides, and polymerizations of activated nucleotides have been reported. Research in all of these areas is ongoing. While we are not yet able to synthesize nucleic acid-like polymers *de novo*, we do have an understanding of the types of chemical processes which must occur in order for the hypothetical pre-RNA World to emerge.

Plausibly prebiotic synthesis of aromatic molecules tend focus on one- and two-ringed compounds, typically purines, pyrimidines, and triazines [36, 39, 40, 102, 104, 107, 135, 136, 189]; however, Hud and Anet proposed that a molecular midwife molecule could be synthesized by known chemistries [188]. Such a molecule would necessarily include a larger stacking surface area than a typical purine. Eschenmoser and

Loewenthal have extrapolated from known syntheses of pyrimidines to multicyclic cofactors such as riboflavin [190]. Characterization of HCN polymers has revealed some units which resemble present-day intercalators, though with more ring nitrogens [115]. Further, various 4,5-diaminopyrimidines have been demonstrated to condense to form tricyclic compounds [191]. It is therefore plausible that larger aromatic molecules could have formed on the prebiotic Earth and served as midwives to assemble the earliest nucleic acids.

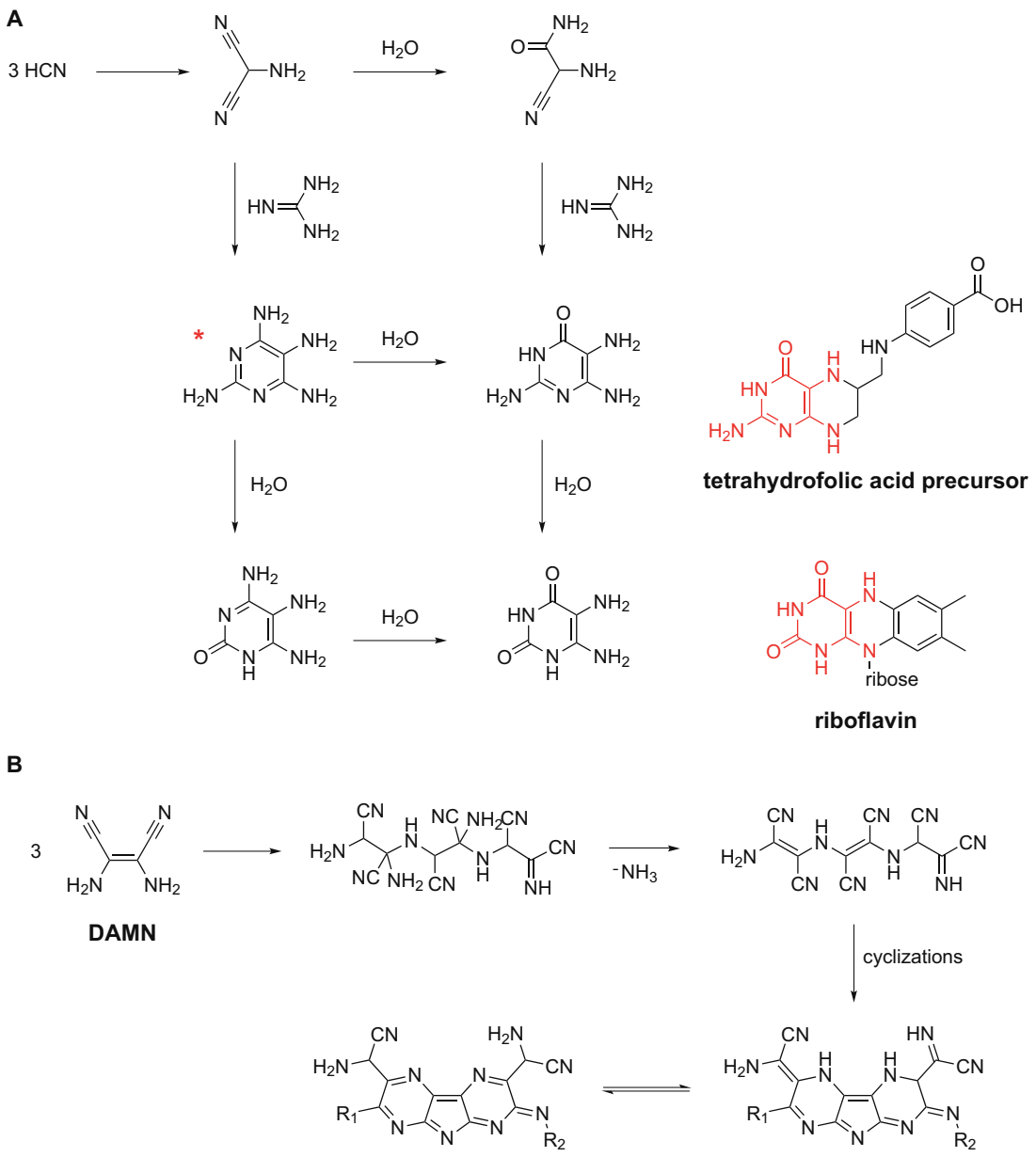

Figure 1.24 Model prebiotic syntheses of larger molecules. (A) Synthesis of tetraaminopyrimidine (indicated by *) and putative water additions result in molecules which could react further to become primitive cofactors, as proposed by Eschenmoser and Loewenthal. Experimental steps, excluding the later water additions, have been verified. The presence of tetrasubstituted pyrimidines in present-day cofactors is suggested by the atoms and bonds pictured in red. Tetrahydrofolic acid may be formed from the illustrated compound by the addition of a glutamic acid residue at the carboxylic acid group, forming an amide bond. Adapted from reference [192]. (B) Synthesis of multi-ringed heterocyclic compounds from DAMN (diaminomaleonitrile, also known as the HCN tetramer) as demonstrated by Ruiz-Bermejo and coworkers. R_1 may be an amine and R_2 may be a proton; alternatively, one or both could be additional heterocyclic units such as those illustrated. Adapted from reference [115].

1.9 WHAT CHEMISTRY TELLS US ABOUT THE PREBIOTIC ENVIRONMENT

Numerous disagreements exist within the scientific community about the correct approach to prebiotic chemistry. Nearly every aspect – from the most basic questions of climate, to the roles of mineral surfaces, to whether compounds could have been synthesized on Earth or were delivered via meteorites or IDPs, to the route to pyrimidine nucleosides and nucleotides – is in dispute to some degree.

Perhaps the best way to move forward in origin of life research is to focus on what the chemistry tells us. We know that life exists on Earth today and that it had to originate by some mechanism or combination of mechanisms. Instead of confining ourselves to scenarios involving a specific atmospheric composition or mineral, we should instead be asking the question: do molecules of biological relevance form under a given set of experimental conditions?

Earth has been subjected to resurfacing since the period of heavy bombardment, and we may never know the precise mineral or atmospheric conditions that existed 3.8 Ga before the present day – though we can make educated guesses [87]. If we can find a set of circumstances under which a required molecule is generated, we can infer that those circumstances may have been relevant at some stage of chemical evolution. Not all molecules of life need be synthesized under the same set of conditions as long as a plausible scenario exists for eventually bringing all of the requisite compounds together.

CHAPTER 2

GUANINE, ADENINE, AND HYPOXANTHINE PRODUCTION IN UV-IRRADIATED FORMAMIDE SOLUTIONS: RELAXATION OF THE REQUIREMENTS FOR PREBIOTIC PURINE NUCLEOBASE FORMATION*

2.1 INTRODUCTION

2.1.1 Information First Versus Metabolism First

The earliest hypotheses about the origin of life centered on metabolic processes, including the concentration by Oparin's coacervates of organic compounds into hydrophobic droplets [193]. Another metabolism-first hypotheses about the origin of life is Wächtershäuser's "iron-sulfur" world which incorporates the oxidative formation of pyrite as an energy source for a primitive, surface-linked metabolic cycle [88-91, 194]. Recent calculations, however, have cast doubt on the validity of a pyrite-pulled metabolism [92].

The best-known "information first" hypothesis is the RNA World (see Section 1.2); this term was coined by Walter Gilbert [28] after the Cech and Altman labs discovered that RNAs could have catalytic activity [26, 27]. Cech has suggested that an RNA World likely evolved from multiple pre-RNA ancestral systems, each capable of self-replication [35].

An information first strategy for the origin of life is attractive for several reasons. Since any pre-RNAs would likely have been capable of reproducing in a semi-conservative manner through base-pairing, we can extrapolate from what we know of

* The results presented in this chapter were previously published in Barks, H.L. *et al*, *ChemBioChem*, **2010**, *11*, 1240-1243.

current nucleic acids to speculate about their prebiotic precursors. For example, stacking interactions are important in the assembly of nucleic acids [145] and suggest a means by which the components of pre-RNAs may have come together in aqueous solutions. Second, many experiments have demonstrated aspects of the construction of nucleic acid chains in enzyme-free solutions (see Section 1.5). The entire process, start-to-finish, has not been accomplished in one pot, and may never be, as the conditions favoring nucleobase synthesis may be different from those favoring polymerization of shorter oligonucleotides into longer ones. One can, however, envision a changing environment such as a tide pool or a geothermal field [195] giving rise to successive conditions conducive to each step in the sequence. Once catalytic pre-RNAs were formed – even inefficient ones – the evolution of functional proteins and of DNA could begin. Conversely, it is difficult to imagine how a primitive metabolic cycle could encode itself in a nucleic acid sequence which would have to produce protein enzymes to catalyze reactions between species participating in that cycle.

If, then, the origin of life proceeded in an information first manner, it is of interest to investigate the means by which the recognition elements of nucleic acids – the nucleobases – first came about. These moieties are key to assembly, via stacking interactions, as well as to recognition through hydrogen bonding.

2.1.2 Previously Reported Syntheses of Purines and Implications

The abiotic production of purines has been the subject of much scientific inquiry over the past fifty years, from Oró's synthesis of adenine from HCN in 1961 [39, 107] (see also Section 1.4.4.2) to the more recent detection of adenine and hypoxanthine in formamide reactions by Saladino and Di Mauro [36]. Additional purines – including

some not found in life – have been detected in carbonaceous chondrites; their synthesis is reported to be consistent with ammonium cyanide chemistry [119].

The abiotic production of adenine is facile; inosine is present in the wobble position of the anticodon loop of some tRNAs. Further, purine bases are ubiquitous in metabolism (e.g. as moieties in cofactors such as *S*-adenosyl methionine and nicotinamide adenine dinucleotide, as well as in coenzyme A). These facts have led scientists to hypothesize that the first nucleic acids were composed of only purine bases, with pyrimidines being incorporated at a later evolutionary stage [24].

2.1.3 Formamide Chemistry

The mechanism of nucleobase production from formamide was first investigated in the 1970s by a team of researchers in Japan; they focused exclusively on unsubstituted purine generated from neat formamide [125, 126]. It was later discovered that the addition of minerals, clays, or salts to formamide solutions altered reaction pathways and produced measurable yields of other compounds of interest.

2.1.3.1 Catalysis by Minerals and Inorganic Salts

In 2001, Saladino and Di Mauro identified several nucleobases in formamide solutions that had been heated in the presence of clays or inorganic salts such as kaolin or calcium carbonate [196]. Since that time, these researchers have examined the catalytic effects of phosphate minerals [82], iron-sulfur minerals [37], titanium dioxide [72], montmorillonites [78], and cosmic dust analogues [129], among other compounds. More recently, Kamaluddin and coworkers reported on formamide condensations catalyzed by iron oxides and iron hydroxides [80]. It has thus been clearly established that different clays, salts, and minerals in combination with formamide generate different product

profiles, both in terms of identities of compounds as well as their yields; unsubstituted purine is ubiquitous in these reactions, but other products seem to rely more heavily on the presence of specific catalysts [36].

2.1.3.2 Effects of UV Irradiation

Although the atmosphere of the early Earth lacked an ultraviolet-blocking ozone layer [197], few studies have investigated the effects of UV irradiation on formamide solutions. Saladino, Di Mauro, and coworkers heated formamide to 160°C in the presence of titanium dioxide; when these reaction solutions were exposed to sunlight, the synthesis of thymine from formamide was reported for the first time [72]. However, this light would have been filtered through the ozone layer, blocking all UV C radiation (100-280 nm) and a large portion of the UV B radiation (280-315 nm) produced by the sun [197]. While solar radiation of UV A (315-400 nm), B, and C reaching the Earth's surface has increased since the period of planetary accretion, prior to the origin of life and the advent of oxygenic photosynthesis, photon flux within the UV B and C ranges would have been higher than at present [197]. Thus, exposure to sunlight is perhaps not the best model for prebiotic ultraviolet irradiation.

Senanayake and Idriss, meanwhile, irradiated formamide on a TiO_2 single crystal surface with 3.2 eV photons (approximately 388 nm) under high vacuum; they were able to observe the formation of all five canonical nucleobases (adenine, thymine, uracil, cytosine, and guanine) found in DNA and RNA [71]. These conditions may be more directly applicable to formamide condensations in the interstellar environment than on the early Earth, and do not account for the higher energy UV C photons that may have

reached the surface of the planet prior to the advent of oxygenic photosynthesis and the formation of an ozone layer.

2.1.4 Analytical Methods for the Separation of Complex Mixtures

Previous analyses of formamide reaction products have primarily been accomplished using gas chromatography coupled with mass spectrometry (GC-MS) [36, 37, 72, 78, 82, 129, 196]. GC-MS can be useful, but suffers from some drawbacks that can be countered by employing complementary techniques. For example, many products of formamide reactions are polar and nonvolatile and must be derivatized to improve peak shape and transit through the GC column. Thus, compounds that do not derivatize well may be underrepresented or undetected in the final analysis of products. Guanine is one such compound.

Another method for identification of formamide reaction products is temperature programmed desorption (TPD) followed by mass spectrometry. This technique, which does not make use of a chromatographic separation method, was employed by Senanayake and Idriss and may not distinguish between nucleobases and their isomers which have identical formulas and masses and similar fragmentation patterns [71].

2.1.4.1 LC-MS/MS

We wished to introduce additional methods of separation and analysis to the investigation of formamide reactions; one of these was liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). LC-MS/MS does not require derivatization of samples, only resuspension in water and dilution into the chromatographic mobile phase. We do note that compounds which are not particularly

soluble in water or in common organic mobile phases may be underrepresented in any LC method.

Since formamide reaction mixtures are known to be complex, one MS/MS method which is particularly useful for their analysis is multiple reaction monitoring, or MRM. MRM is performed on a triple quadrupole mass spectrometer (Figure 2.1) after LC separation is accomplished. Mobile phase exits the LC column and is ionized. Ions of a target mass representing an analyte of interest, known as precursor ions, are selected in the first quadrupole. The second quadrupole is a chamber where the precursor ions are fragmented via collisions with argon. In the third quadrupole, product ions are selected; compounds reach a detector after exiting the third quadrupole. Products are identified by chromatographic retention time, precursor ion mass, and product ion mass.

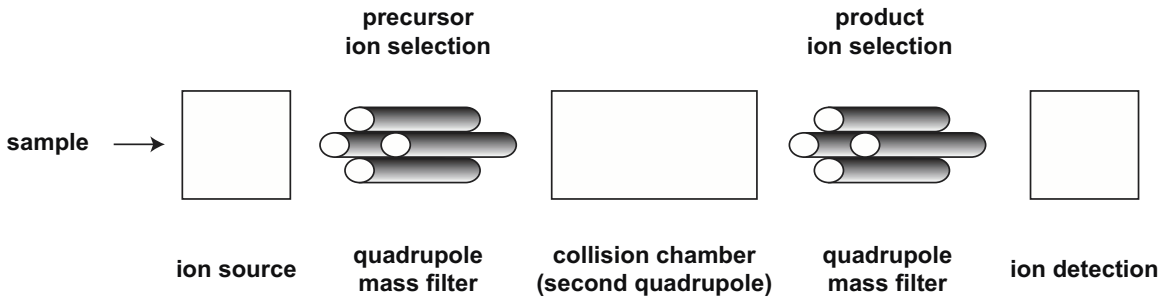

Figure 2.1 Schematic of a triple quadrupole mass spectrometer.

MRM allows analytical chemists to achieve baseline-level resolution of peaks, even under separation conditions which are not ideal. A total ion chromatogram with many overlapping peaks may thus be greatly simplified, allowing quantitation of selected products (Figure 2.2).

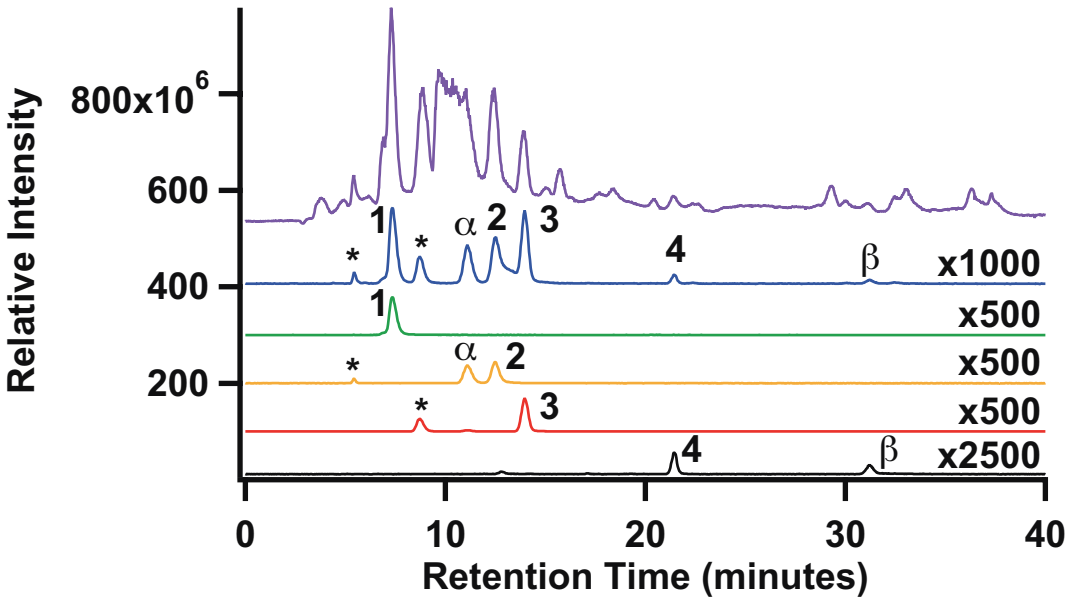

Figure 2.2 Multiple reaction monitoring allows baseline-level resolution of peaks. Purple chromatogram is the total ion chromatogram (m/z 50 to 300). Blue is the combined selected ion chromatogram (numerical sum of green, orange, red, and black traces). The green trace displays the 120.9 → 93.9 transition, orange the 135.9 → 118.7 transition, red the 136.9 → 109.8 transition, and black the 151.8 → 134.8 transition. * represents any unidentified reaction product which shares a precursor and product ion with known standard, but which has a different retention time. **1** purine, **2** adenine, **3** hypoxanthine, **4** guanine, α internal standard 2-aminopurine, β internal standard isoguanine.

One drawback of MRM is that only known products can be quantified. When setting up MRM parameters, a user must optimize cone and collision voltages for the detection of specific product and precursor ions. Further, when working with many related compounds – as happens with formamide reactions – unidentified peaks which represent isomers of target analytes may still appear in selected ion channels (Figure 2.2, peaks marked with *). Finally, as with any LC-MS method, internal standards are required to control for variations in ionization among injections, and calibration curves must still be constructed to correct for differential ionization of standards.

2.1.4.2 High Pressure Liquid Chromatography

HPLC, or high pressure liquid chromatography (sometimes called high performance liquid chromatography), is a second LC technique we applied to the analysis of formamide reaction mixtures. Like LC-MS/MS, HPLC requires no sample derivatization, but also like LC-MS/MS, HPLC may not detect compounds which are insoluble in water or common organic mobile phases.

A variety of detectors may be used with HPLC; we chose a diode array detector (DAD) which allows the simultaneous monitoring of absorbance of multiple wavelengths of light by analytes. A full UV-Vis spectrum for each chromatographic peak can be obtained in this manner, though analytes are limited to molecules containing chromophores. Because purine nucleobases contain chromophores, HPLC coupled with a DAD is ideally suited to the analysis of formamide reaction mixtures.

HPLC's primary advantage over LC-MS/MS is that the user is not limited to the analysis of compounds for which authentic standards are possessed; further, ionization and fragmentation voltages do not need to be optimized. UV-Vis detection with a DAD is somewhat less sensitive than LC-MS/MS detection, however, and minor products may be missed with conventional HPLC. HPLC and LC-MS/MS are complementary when used in tandem, however, and allow the chemist to identify and quantify even minor products while still retaining a sense of the complexity of the reaction mixture.

2.1.5 Nuclear Magnetic Resonance Spectroscopy for the Determination of Bulk Solution Conditions

Nuclear magnetic resonance (NMR) spectroscopy is best applied to high-concentration samples of pure compounds. ^{1}H NMR spectra of formamide reaction

mixtures dried under vacuum and resuspended in deuterated solvents are marked by many overlapping resonance peaks, which can be difficult to interpret. However, NMR is able to provide information about bulk solution conditions. The efforts of scientists to obtain "neat" formamide have long been frustrated by the hydrolysis of formamide to HCN and water, and the reaction of water with formamide to generate ammonium formate [117]. Since ammonium formate has been implicated in the synthesis of adenine from DAMN (diaminomaleonitrile, the HCN tetramer) [116], the amount of ammonium formate present in formamide reaction solutions is of interest.

The ammonium ion is difficult to distinguish by ^{1}H NMR from other amines and amides in formamide solutions, but the formate ion possesses a proton resonance slightly downfield from that of the nonexchangeable (amide) formamide proton. Formate is generated in sufficient amounts by heating formamide for a short time that it can be easily quantified relative to formamide using ^{1}H NMR.

2.2 EXPERIMENTAL PROCEDURES

2.2.1 Materials

Solvents were HPLC grade or LC-MS grade, as appropriate, and were purchased from EMD chemicals. Inorganic catalysts, HPLC buffers, and formamide were purchased from Sigma Aldrich, Fisher Scientific, or VWR, and were used as received. Nucleobase standards and DAMN were purchased from Sigma Aldrich, AICN from MP Biomedical, and AICA from TCI America. Water was purified using a Barnstead nanoPure system and all HPLC buffers were filtered with 0.2 μ solvent filters (Millipore) prior to use.

2.2.2 Reaction Conditions

Reactions were carried out in quartz test tubes in a custom-built isothermal aluminum heating block with temperature controlled to within ±0.5°C by microprocessor. Reactions were stirred with Teflon-coated stir bars and test tubes were covered with glass to prevent evaporation. Two of the test tube wells in the heating block were slotted to allow exposure to a low pressure mercury lamp (Pen-Ray) with primary emission at 254 nm and no emission of photons with wavelengths below 200 nm. All reaction solutions were prepared immediately prior to placement in the thermally-equilibrated reaction center.

Most reaction solutions were heated for 48 h at 130°C with 200 mg of catalyst (when catalyst was used) in 4 mL of formamide. In some experiments, a small molecule intermediate was spiked into neat formamide at a concentration of 10 mM. These reactions were heated for 96 h, with aliquots removed for analysis over time.

2.2.3 Sample Preparation

For HPLC and LC-MS/MS analysis, 200 μL aliquots of crude reaction mixtures were dried under vacuum at 60°C prior to injection, as formamide interfered significantly with chromatography.

2.2.3.1 LC-MS/MS

For LC-MS/MS, dried aliquots of formamide reaction mixtures were resuspended in 200 μL water, agitated with a vortex mixer, and centrifuged to pellet insoluble products. 5 μL aliquots of the supernatants of these centrifuged samples were added to 285 μL of acetonitrile and 10 μL of a solution containing the internal standards 2-aminopurine and isoguanine (Figure 2.3). For quantitation purposes, stock solutions of

authentic standards of the nucleobases purine, adenine, hypoxanthine, and guanine (Figure 2.3) were also prepared. At least five different concentrations of each, in the same 95:5 acetonitrile:water mixture as for reaction samples, were injected into the LC for the purpose of establishing calibration curves. Spiking of authentic standards into reaction mixtures was also performed to confirm product identifications.

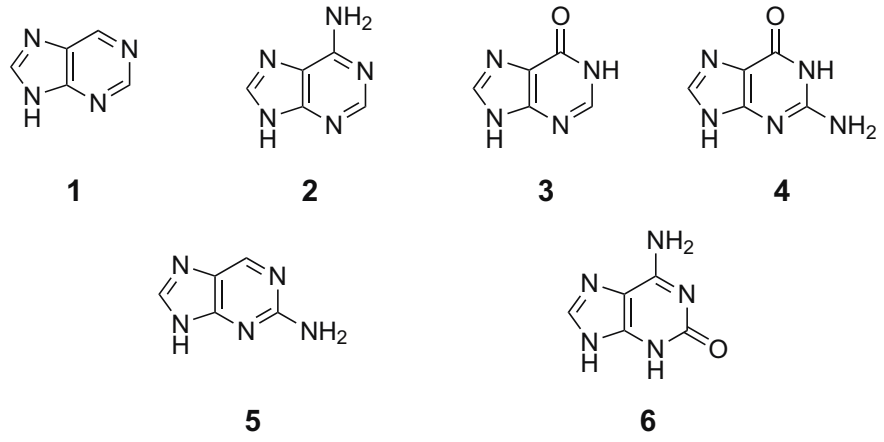

Figure 2.3 Nucleobase standards used in this study. **1** purine, **2** adenine, **3** hypoxanthine, **4** guanine, **5** 2-aminopurine, **6** isoguanine.

2.2.3.2 HPLC

For HPLC analysis, dried aliquots of formamide reaction mixtures were resuspended in 500 μL water, agitated with a vortex mixer, and centrifuged to pellet insoluble products. Supernatant aliquots were diluted and in some cases, spiked with authentic standards prior to injection. For quantitation purposes, stock solutions of the nucleobases purine, adenine, and hypoxanthine were also prepared. At least five different concentrations of each were injected into the LC for the purpose of establishing

calibration curves. Guanine could not be observed in the HPLC analyses due to its low solubility in water.

2.2.3.3 NMR

NMR analyses were performed on crude reaction mixtures; a flame-sealed capillary insert containing a TMS (tetramethylsilane) was included as a standard. The TMS chemical shift was set to 0 ppm.

2.2.4 Instrumental Methods

2.2.4.1 LC-MS/MS

2.2.4.1.1 Liquid Chromatography

LC separations of formamide reaction mixtures were achieved on an Agilent 1100 binary HPLC system. A Merck SeQuant ZIC-HILIC column (PEEK, 100 × 2.1 mm, 5 μm particle size, 200 Å porosity) was used for separations. The column was maintained at 30°C with a flow rate of 0.1 mL/min. The injection volume was 30 μL.

Solvent A was acetonitrile; solvent B was 2.5 mM ammonium formate and 25 mM formic acid in water. The following gradient was used: 0-20 minutes, 5-20% B; 20-30 minutes, 20% B; 30-31 minutes, 20-5% B; 31-40 minutes, 5% B.

2.2.4.1.2 Multiple Reaction Monitoring

MRM was performed on a Micromass Quattro LC triple quadrupole mass spectrometer (see Table 2.1 for optimal MRM parameters). The source block temperature was 100°C and the desolvation temperature was 150°C. The dwell time for each MRM transition was 0.2 seconds and the interchannel delay was 1.8 seconds.

Table 2.1 LC-MS/MS parameters used for optimal detection of nucleobase standards.

Nucleobase Standard	Precursor Ion Mass (Da)	Product Ion Mass (Da)	Collision Voltage (V)	Cone Voltage (V)	Retention Time (minutes)
Purine	120.9	93.9	20	40	7.6
Adenine	135.9	118.7	35	50	12.4
Hypoxanthine	136.9	109.8	30	50	14.1
Guanine	151.8	134.8	25	40	21.5
2-Aminopurine	135.9	118.7	35	50	11.1
Isoguanine	151.8	134.8	25	40	30.6
AICA	126.9	54.8	50	20	15.1
AICN	108.8	81.8	20	20	5.7
DAMN	108.8	81.8	10	20	6.2

2.2.4.2 HPLC

HPLC analysis was performed on an Agilent 1200 RRLC system equipped with a DAD. Separations were reverse-phase and accomplished using a Phenomenex Synergi Polar RP column (250 × 4.6 mm, 4 μm particle size). The column was maintained at 30°C with a flow rate of 0.75 mL/min.

Solvent A was 20 mM sodium phosphate buffer at pH 7; solvent B was 10% acetonitrile in methanol (v/v). The following gradient was employed to achieve separation: 0-12 minutes, 5-10% B; 12-16 minutes, 10-20% B; 16-18 minutes, 20-40% B; 18-19 minutes, 40-5% B; 19-30 minutes, 5% B.

2.2.4.3 NMR

^{1}H NMR spectroscopy was performed on a 400 MHz Varian Mercury Vx instrument. Spectra were acquired unlocked as single transients; sixteen individual spectra were averaged after an included TMS standard was set to 0 ppm.

2.3 RESULTS

2.3.1 Thermal Reactions

When neat formamide was heated at 130°C for 48 h, the only nucleobase product generated was purine, a result consistent with experiments conducted at higher temperatures [36] (Figure 2.4). HPLC monitored at 260 nm revealed the presence of additional unidentified peaks which did not share retention times with purine or pyrimidine standards. The addition of calcium carbonate enhanced the yield of purine and also generated small amounts of adenine and hypoxanthine as evidenced by both LC-MS/MS and HPLC analysis (Figure 2.4). When sodium pyrophosphate was added to reaction mixtures, purine yield was significantly enhanced, as were yields of adenine and hypoxanthine, and the production of a small amount of guanine was observed (Figure 2.4 A, inset).

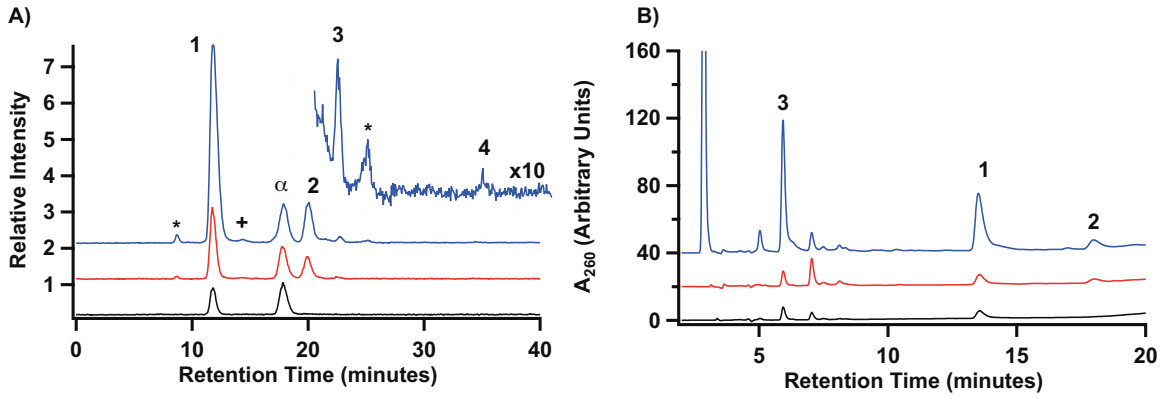

Figure 2.4 Formamide reactions conducted at 130°C for 48 h. See Section 2.2.2 for experimental details. Black traces represent products of "neat" formamide reactions, red represent formamide with $CaCO_3$ catalyst, and blue represent formamide with $Na_4P_2O_7$ catalyst. (A) LC-MS/MS combined selected ion chromatograms; (B) HPLC chromatograms. * peak has same product and precursor ion masses as adenine but a different retention time, + peak has same product and precursor ion masses as hypoxanthine but a different retention time, **1** purine, **2** adenine, **3** hypoxanthine, **4** guanine, **a** LC-MS/MS internal standard 2-aminopurine. Note that hypoxanthine co-elutes with an unidentified product under our HPLC conditions.

Our thermal-only formamide reactions at 130°C thus display many similarities to previous work conducted at 160°C, with different mineral and/or inorganic catalysts producing unique profiles of products with varying yields.

2.3.2 Thermal and Photochemical Reactions

Irradiation of formamide reaction solutions with 254 nm ultraviolet light, meanwhile, altered yields and increased product diversity, regardless of whether a catalyst was present, and regardless of which catalyst was present (Figure 2.5). While relative yields of the various products were dissimilar, identities of products were the same, with notable increases in the amounts of hypoxanthine and guanine produced (Figure 2.5).

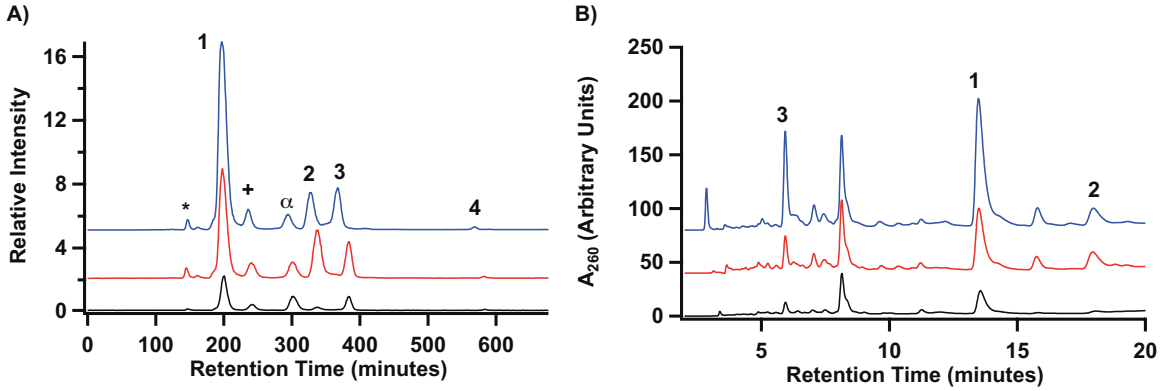

Figure 2.5 Formamide reactions conducted at 130°C for 48 h with 254 nm UV irradiation. See Section 2.2.2 for experimental details. Black traces represent products of "neat" formamide reactions, red represent formamide with $CaCO_3$ catalyst, and blue represent formamide with $Na_4P_2O_7$ catalyst. (A) LC-MS/MS combined selected ion chromatograms; (B) HPLC chromatograms. * peak has same product and precursor ion masses as adenine but a different retention time, + peak has same product and precursor ion masses as hypoxanthine but a different retention time, **1** purine, **2** adenine, **3** hypoxanthine, **4** guanine, **a** LC-MS/MS internal standard 2-aminopurine. Note that hypoxanthine co-elutes with an unidentified product under our HPLC conditions.

2.3.3 HCN Reaction Pathways

Most published abiotic syntheses of purines involve HCN chemistry (see Sections 1.4.4.2 and 2.1.2), and formamide is known to equilibrate with HCN and water (see Section 1.4.4.1). We thus wanted to investigate the possibility that well-characterized HCN chemical pathways were active in formamide. We were unable to detect key intermediate compounds (DAMN, AICA, and AICN) by HPLC or LC-MS/MS. It is possible that these compounds were generated *in situ* and rapidly underwent further reactions including the production of nucleobases. We investigated aspects of this hypothesis by spiking DAMN, AICA, and AICN into formamide and monitoring the production of adenine and hypoxanthine over time.

AICN and DAMN (Figure 2.8) are hypothesized precursors to adenine; both have the formula $C_4H_4N_4$. Whereas DAMN requires isomerization about a double bond, followed by cyclization, to generate adenine, AICN has already undergone this cyclization. UV light is one possible source of the energy required for DAMN isomerization, so in the absence of UV irradiation, we expected that AICN solutions would produce adenine but that DAMN solutions would be unreactive. This hypothesis was confirmed; over a 96 h period, aliquots of AICN-spiked reactions generated increasing amounts of adenine, while spiking of other compounds, as well as a neat formamide control, generated none (Figure 2.6 A). UV irradiation of identical solutions still resulted in the production of adenine in AICN-spiked reactions, but now DAMN-spiked mixtures also generated this nucleobase (Figure 2.6 B). A leveling-off of adenine levels under conditions of UV irradiation suggests a steady state of production and destruction of this product (Figure 2.6 B).

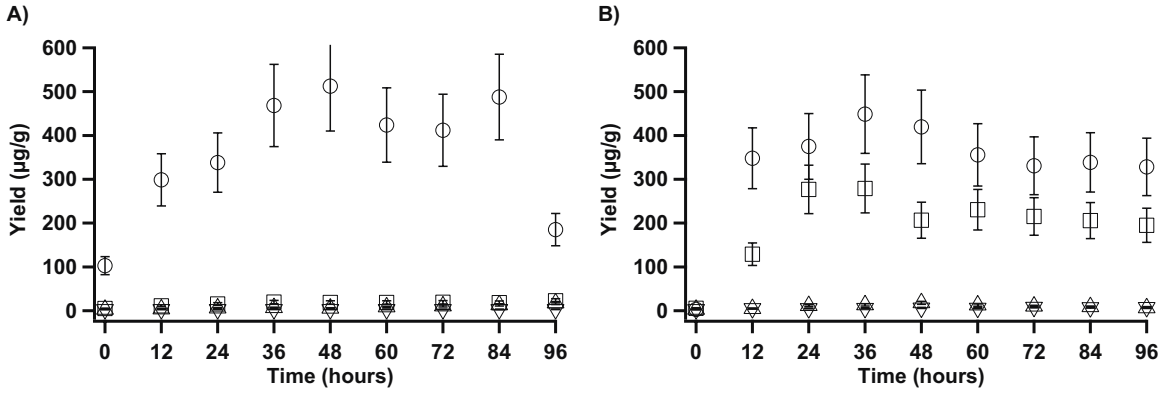

Figure 2.6 Yields of adenine over time (reported in µg/g formamide) as determined by LC-MS/MS for (A) thermal-only reactions and (B) thermal and photochemical reactions. (▽) represents "neat" formamide, (□) 10 mM DAMN in formamide, (△) 10 mM AICA in formamide, and (○) 10 mM AICN in formamide. See Section 2.2.2 for experimental details.

Whereas AICN and DAMN are believed to be precursors to adenine, AICA is a reported hypoxanthine precursor. Like AICN, AICA is cyclized, and we hypothesized that hypoxanthine production would occur even under thermal-only reaction conditions. This was confirmed by LC-MS/MS analysis; hypoxanthine yields remained fairly steady over the course of a 96 h reaction (Figure 2.7 A). UV irradiation complicated the situation; AICA solutions still had an initial advantage in hypoxanthine production, but at the end of the reaction period, all solutions (unspiked, or spiked with AICA, AICN, or DAMN) yielded approximately 100 µg hypoxanthine per gram of formamide reacted (Figure 2.7 B). These results suggest that some photodegradation of hypoxanthine may occur, but also that the various pathways to hypoxanthine production are active under combined thermal and photochemical conditions.

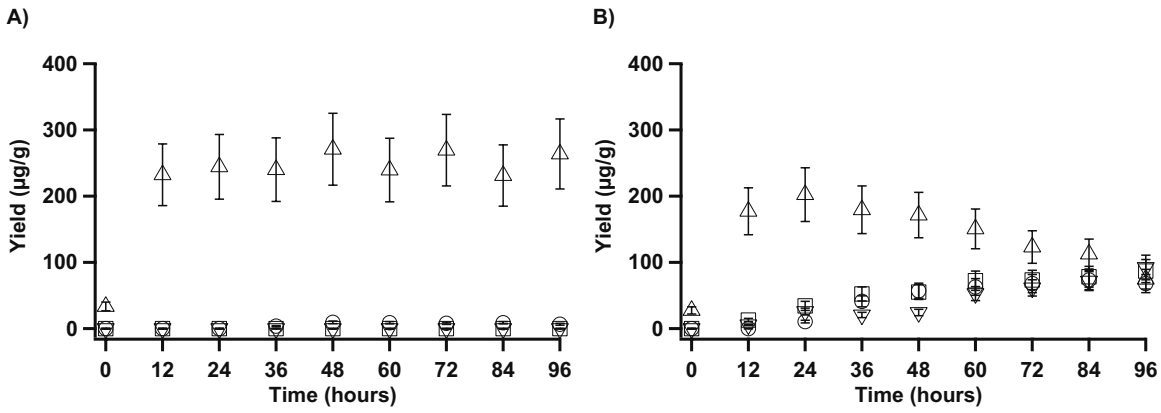

Figure 2.7 Yields of hypoxanthine over time (reported in µg/g formamide) as determined by LC-MS/MS for (A) thermal-only reactions and (B) thermal and photochemical reactions. (▽) represents "neat" formamide, (□) 10 mM DAMN in formamide, (△) 10 mM AICA in formamide, and (○) 10 mM AICN in formamide. See Section 2.2.2 for experimental details.

Our results are consistent with previously published pathways of formation of nucleobases from formamide. While theoretical calculations now confirm that multiple pathways of formamide decomposition compete with one another [198], the earliest steps of the scheme outlined by Duvernay and colleagues in 2005 [199] could be key in generating the HCN required for the condensation reactions that generate DAMN and other HCN intermediates. A summary of previously-published and proposed pathways, as well as our results from spiking interactions, is presented in Figure 2.8.

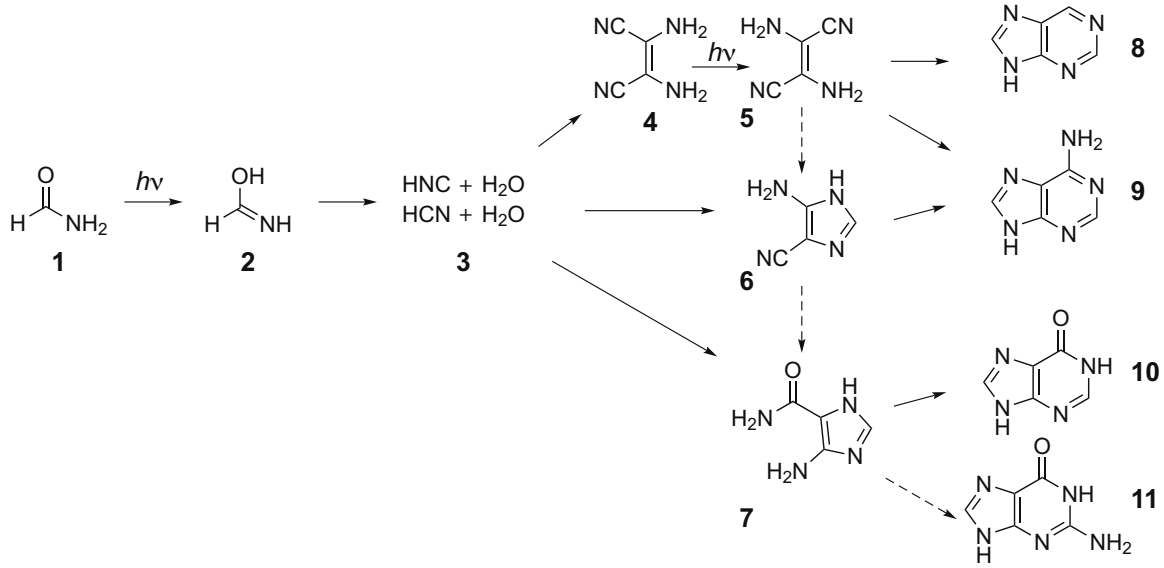

Figure 2.8 Putative pathway from formamide to purines. Dashed arrows represent unconfirmed steps. **1** formamide, **2** formamidic acid, **3** hydrogen cyanide and hydrogen isocyanide, **4** DAMN (diaminomaleonitrile), **5** DAFN (diaminofumaronitrile), **6** AICN (aminoimidazolecarbonitrile), **7** AICA (aminoimidazolecarboxamide), **8** purine, **9** adenine, **10** hypoxanthine, **11** guanine.

2.3.4 Mixed Water-Formamide Solutions

We have chosen to work with neat formamide solutions for practical experimental considerations, including the ability to speed reactions by working at temperatures above the boiling point of water. It is unlikely, however, that pools of neat formamide existed on the early Earth; water was almost certainly present [200, 201] and more abundant than formamide. Since water and formamide are miscible without forming an azeotrope [118], and since formamide boils at a higher temperature than does water, we envisioned a "drying pool" scenario in which water could evaporate from mixed formamide/water solutions, concentrating formamide during hot periods and enabling reactions which generate nucleobases.

We performed an experiment wherein we began with a 90:10 solution of water and formamide (mol/mol) and heated it at a lower temperature (100°C) for 96 h. Water did evaporate from this reaction mixture, and some nucleobase products were identified, including purine, adenine, and hypoxanthine (Figure 2.9).

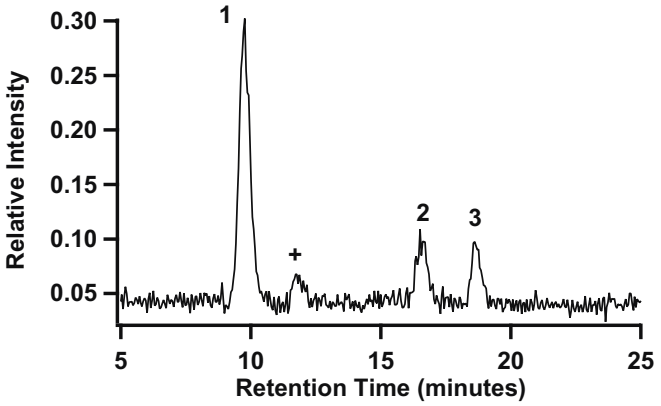

Figure 2.9 Combined selected ion chromatogram for the reaction of 10 mol% formamide in 90 mol% water for 96 h at 100°C. **1** purine, + product with same mass as hypoxanthine but different retention time, **2** adenine, **3** hypoxanthine.

This result is significant not only because it is consistent with our drying pool hypothesis, but because two of the products generated – adenine and hypoxanthine – have not been identified as products of thermal-only reactions of neat formamide; in this case, water seems to have been necessary for their production. The observed yields of all products were lower than when starting from water-free systems, but this is not surprising given the lower temperature and lower concentrations of formamide at the beginning of the reaction.

2.3.5 Bulk Solution Conditions

In addition to the loss of water noted in our drying pool experiment, we were interested in the rate of hydrolysis of formamide to ammonium formate, since formamide has been historically difficult to purify [117] and since solutions of ammonium formate and DAMN have been demonstrated to produce adenine [116]. While no formate is detectable by NMR in neat formamide prior to heating, after 24 hours, approximately 1% conversion occurs (Figure 2.10). The presence of ammonium formate at these concentrations does not appear to interfere with the genesis of purine nucleobases; it is possible that *in situ* production of ammonium formate is actually an important factor in nucleobase production.

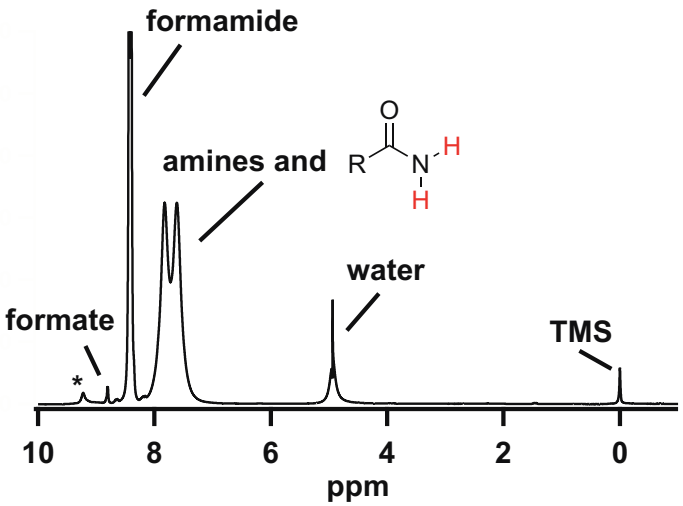

Figure 2.10 ^{1}H NMR spectrum of formamide heated for 24 h at 130°C. The peak labeled * is an unidentified reaction product; the overlapping resonances at 7.5-8 ppm represent all amine protons as well as protons associated with amide nitrogens (in red). TMS = tetramethylsilane. The spectrum is enlarged to highlight smaller peaks near the baseline; the full height of the formamide peak is not shown.

2.4 DISCUSSION

We have demonstrated for the first time that nucleobases other than unsubstituted purine are produced in heated, UV-irradiated formamide solutions, in the absence of a mineral or inorganic salt. This work also represents the first identification of guanine from a formamide reaction solution.

2.4.1 Relaxation of the Requirements for Prebiotic Purine Nucleobase Production

Our work also represents an advance in the field of prebiotic formamide chemistry in that we are able to generate nucleobase products under a variety of relaxed conditions.

2.4.1.1 Reduced Temperature

Early work to determine the mechanism of formation of purine from formamide was conducted over a range of temperatures from 150°C to 200°C [125, 126]. Later reactions performed by Saladino and Di Mauro were conducted at 160°C [36]. Most of our reactions were conducted at 130°C, which represents a significant relaxation of temperature requirements for the production of nucleobases. Further, we have demonstrated that some purine bases are produced even at 100°C if the reaction is allowed to proceed for a longer period of time.

2.4.1.2 UV Irradiation Decreases Dependence on Specific Catalyst

While the presence of an inorganic catalyst affects the absolute and relative yields of purine nucleobase products from formamide solutions, our work demonstrates that UV irradiation reduces reliance on specific catalysts (present as solids, dissolved salts, or a combination of the two) in these reactions.

Our results indicate that formamide, or one of its thermal products, is a photoactive species. While the VUV absorption spectrum of formamide nominally ends at 206 nm (photon energy 6.03 eV) [202], this molecule possesses an absorbance tail extending past 254 nm (ε_{254} = 0.6 $M^{-1} \cdot cm^{-1}$, photon energy 4.89 eV). The photochemical steps leading from formamide to HCN and water have been outlined by Duvernay and colleagues [199] (see also Figure 2.8). In the likely scenario that optical excitation of formamide is a photochemically important transition in our reactions, this thermal or UV-promoted generation of HCN from formamide provides a source of water even in neat formamide solutions; this production of water may be vital to further steps in these reaction pathways.

2.4.2 Support for HCN Condensation Pathways Being Active in Formamide Solutions

HCN condensation to form adenine was the earliest reported plausibly prebiotic synthesis of a nucleobase [39, 107]; in the intervening years, DAMN, AICA, and AICN were established as important intermediates in the pathway to adenine and other purines [120, 121, 123, 124, 203]. While AICA has been detected as a product in formamide reactions by GC-MS [36], we were not able to identify this compound in our reaction mixtures. Nonetheless, in experiments where DAMN, AICA, and AICN were added to formamide solutions prior to heating, we observed increased yields of expected products: adenine from AICN and hypoxanthine from AICA in thermal-only reactions, and adenine from DAMN in combined thermal and photochemical reactions. These results are consistent with the activity of well-established cyanide condensation pathways in formamide solutions.

2.4.3 The "Drying Pool" Scenario

As discussed previously, it is unrealistic to assume that pools of neat formamide existed on the prebiotic Earth. Water has been detected in comets [204] and may have been delivered to the planet's surface by impacts; other hypotheses posit that the Earth accreted in a wet state and that water was endogenous to the planet [205]. Regardless of the source, geological and isotopic evidence places liquid water on Earth at an early stage of planetary development [200, 201].

Our results demonstrate that purine nucleobases are still able to form from a mixed water/formamide solution heated at 100°C for four days. Over the course of this reaction, water evaporated, leaving a more concentrated solution of formamide. Neat formamide, in which nucleobase condensations can occur, could thus have been generated in drying pools through the evaporation of water.

2.4.4 New Analytical Techniques

Through the application of new analytical techniques such as HPLC and LC-MS/MS, we were able to demonstrate the production of guanine from UV-irradiated formamide solutions for the first time. Liquid chromatography techniques should be thought of as complementary to gas chromatography. While LC has certain advantages, including direct analysis of reaction mixtures without derivatization, GC-MS may be better able to separate and detect less polar, more volatile compounds.

2.5 CONCLUDING REMARKS

We have shown that the requirements for prebiotic purine nucleobase production are not as strict as once believed; these compounds can be formed at lower temperatures than previously reported, they can be formed from mixed water-formamide solutions, and

in the presence of 254 nm UV irradiation, they can be generated in the absence of a mineral or inorganic catalyst.

Despite the advances reported here, several questions remain. The inorganic salts used as catalysts in these reactions, sodium pyrophosphate and calcium carbonate, include Group 1 and 2 cations which are known to exist in only one charged oxidation state each (+1 and +2, respectively). It remains to be seen whether photoactive cations capable of changing oxidation states and catalyzing redox reactions – including those of transition metals – produce different results.

Further, both salts used in this study are at least partially soluble in water. Our work does not differentiate between the catalytic activities of ions sequestered in minerals and ions in solution. The roles of insoluble minerals and salts should be investigated to better distinguish between the effects of solid and solution-phase chemistries.

Additionally, we have investigated the influences of photons of only one wavelength (254 nm) on formamide chemistry. Multiple wavelengths of light would have been available on the early Earth, and further experiments should be conducted to determine whether photons of other wavelengths provide sufficient energetic input to stimulate the production of nucleobases from heated formamide solutions.

CHAPTER 3

SYNERGETIC EFFECTS OF PYRITE, UV IRRADIATION, AND ATMOSPHERIC GASES ON THE GENERATION OF NITROGEN HETEROCYCLES FROM FORMAMIDE SOLUTIONS

3.1 INTRODUCTION

Following a bottom-up approach to the origin of the RNA World, one might first ask how the individual components of pre-RNAs were synthesized on a prebiotic Earth. Formamide is attractive as a precursor to the molecules of life because it is simple and stable and contains the essential elements carbon, hydrogen, nitrogen, and oxygen. Formamide exhibits a low volatility compared to either water or HCN, and it is liquid over a wide range of temperatures (from 2°C to 210°C). As described previously, formamide displays a versatile reactivity upon heating in the presence of various mineral catalysts [36, 37]. Further, formamide has been detected in the interstellar medium [206] and in comets [204] and is believed to have been available on the early Earth either via the impact of extraterrestrial materials or as a product of the hydrolysis of HCN. Because formamide does not form an azeotrope with water [118], its concentration from dilute aqueous solutions can be accomplished by gentle heating [40].

We have previously demonstrated that the irradiation of formamide reaction solutions with 254 nm UV light increases the yield and diversity of purine nucleobases produced, even in the absence of a catalyst (see Chapter 2 or reference [40]). The inorganic salts employed as catalysts in our previous work were somewhat soluble in formamide and we were interested in further separating out the effects of solution-phase

formamide chemistry versus chemistry that occurs on a mineral surface. We were also interested in conducting reactions under conditions that more closely resembled those on the primitive Earth. Since all formamide chemistry to date has been carried out under an air atmosphere [36, 40] or in high vacuum [71], we wanted to characterize the products of formamide reactions conducted under both inert and plausibly prebiotic atmospheres.

3.1.1 Mineralogy of the Primitive Earth

Numerous minerals have been invoked in the field of prebiotic chemistry; these include borates [207], clays [79, 85, 208], and a host of phosphorus-containing species of both terrestrial and extraterrestrial origin [82, 93]. When evaluating research in this area, it is important to consider which mineral species may actually have been present at and before the time life began.

3.1.1.1 Elemental Composition

Today, the Earth is composed primarily of iron, oxygen, silicon, magnesium, and sulfur, with smaller percentages of a few additional metals and numerous trace elements [209]. Any mineralogical consideration of the origin of life ought to consider the roles of these elements, as well as the planet's geological and climatological history. Iron oxides are common components of banded iron formations, but most of these were deposited during the Proterozoic – after the origin of life [57, 210, 211]. Some oxygen-containing minerals such as silicates, however, would likely have been present on the early planet [87].

3.1.1.2 Early Planetary Minerals

While pre-stellar clouds already contained some carbides, silicates, nitrides, and oxides, it was not until Earth accreted from its protoplanetary disk that minerals would

have truly begun to diversify through various geologic processes [87]. Aqueous and thermal alteration of early rocks would have played significant roles in mineral evolution, and some minerals may have been delivered from exogenous sources such as meteorite impacts [87]. A simplified timeline of mineral evolution on the primitive Earth is provided in Table 3.1.

Table 3.1 A selection of minerals believed to have been present on the early Earth. Adapted from reference [87].

Age (Ga)	Mineral(s)
>4.56	olivine, pyroxene, iron sulfide
4.56-4.55	sulfates, carbonates, transition metal sulfides and phosphates
4.55-4.0	hydroxides, clay minerals
4.0-3.5	micas, quartz, feldspar, zircon

3.1.1.3 Iron Sulfides and Formamide Chemistry

Pyrite is an iron sulfide mineral of the formula FeS_2 (Figure 3.1) which is widely distributed on Earth today [212] and which is also believed to have been available on a prebiotic Earth [87, 95].

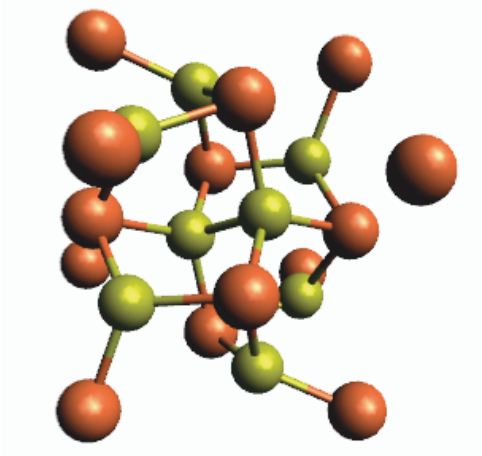

Figure 3.1 Pyrite crystal structure unit cell. Iron ions are pictured in brown and sulfur ions in yellow. Crystal coordinates provided in reference [213].

Pyrite has long been implicated in "metabolism-first" scenarios for the origin of life [89, 91], but its ability to catalyze the formation of nucleobases from formamide has also been examined [37] (Figure 3.2).

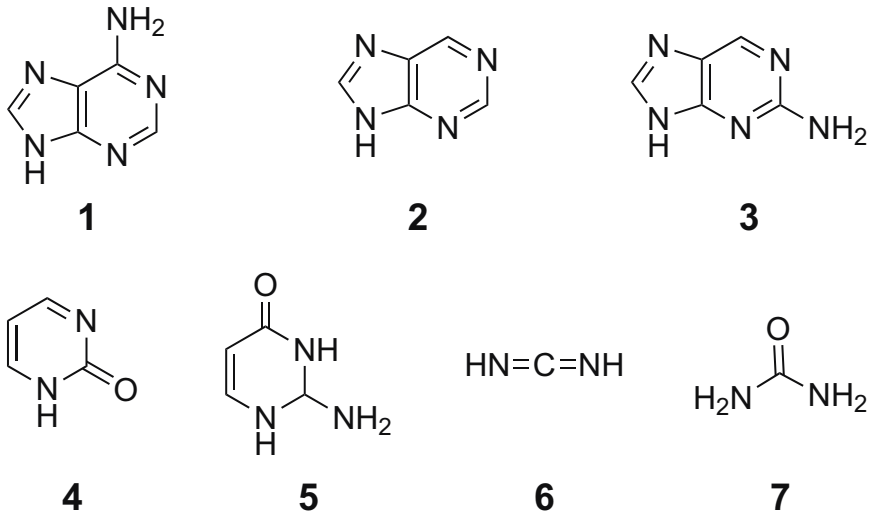

Figure 3.2 Condensation products generated when formamide is heated at 160°C in the presence of pyrite. The same products are generated in similar yields when a non-mineral FeS_2 salt is used as a catalyst [37]. **1** adenine, **2** purine, **3** 2-aminopurine, **4** 2-pyrimidinone, **5** isocytosine, **6** carbodiimide, **7** urea.

Pyrite is known to be more electrochemically inert than other common sulfide minerals, but trace inclusions of other atoms can alter this mineral's electrical conductivity [214]. Iron, which is nearly always present in its ferrous form in pyrite in a 1:2 ratio with sulfur [214], is a transition metal with a wide range of available oxidation states. While pyrite is not particularly soluble in formamide, and may act as a photocatalyst in its mineral form, any iron that escapes into solution could interact with UV photons and undergo changes in oxidation state which are inaccessible to the sodium and calcium cations previously examined as catalysts for formamide chemistry with UV light [40] (see also Chapter 2).

3.1.2 Atmosphere of the Primitive Earth

It is widely accepted that prior to the origin of photosynthetic organisms, Earth's atmosphere contained no oxygen and hence no ultraviolet-blocking ozone layer [57]. However, reasonable and well-informed researchers often disagree on the precise composition of the planet's atmosphere at the time of the dawn of life, basing their conclusions on everything from mineralogy [59] to the Earth's likely temperature and albedo [106, 215].

3.1.2.1 Early Experiments

Perhaps the best-known experiment in prebiotic chemistry was conducted under a reducing atmosphere of methane, ammonia, water vapor, and hydrogen gas [55]. Oparin himself postulated a reducing atmosphere which preceded and which may have inspired the Miller-Urey experiments [50, 51, 54].

3.1.2.2 Current Hypotheses

More recently, multiple models of the prebiotic atmosphere have been proposed, and hypotheses and experimental evidence have been offered in favor of each. These include, but are not limited to, atmospheres rich in carbon dioxide [60], Titan-like hazes of nitrogen and methane [62], and CO_2 atmospheres containing a large percentage of H_2 [63]. A number of trace atmospheric components may also have been available, such as sulfur-containing species from volcanic outgassing [216].

3.1.3 Mechanism of Purine and Adenine Formation from Formamide

Several mechanisms for purine and adenine synthesis from formamide and/or HCN, along with various means of experimental support for these hypotheses, have been published in the last fifty years [39, 40, 126]. More recently, it has been demonstrated that a single, unified mechanism may exist to explain both adenine and purine production from formamide solutions and that the relative yields of these two products may be influenced by the ratio of cyanide ions to formate ions [217] (Figure 3.3).

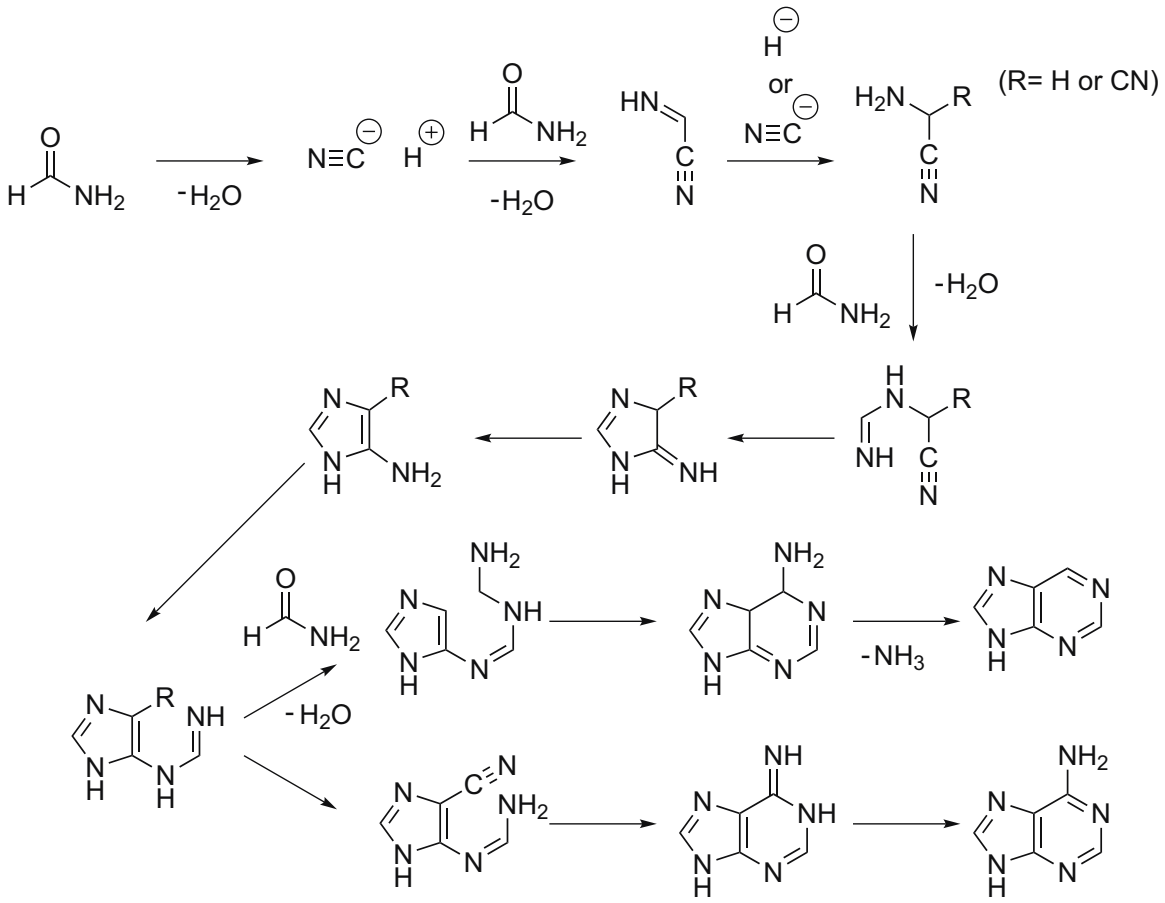

Figure 3.3 Unified mechanism of purine and adenine formation from formamide. Adapted from reference [217].

Since the non-exchangeable formate proton has a chemical shift distinct from its counterpart in formamide, it is possible to measure the conversion of formamide to formate via ^{1}H NMR; further, adenine and purine yields from formamide solutions are easy to measure via HPLC or LC-MS/MS [40]. Additionally, a variety of assays for cyanide determination exist [218-222]. We were interested in whether the unified mechanism of purine and adenine formation was active in mineral-catalyzed, UV-irradiated formamide solutions; we performed tests for cyanide and formate in our reaction solutions and quantified the amounts of purine and adenine produced under

various conditions to determine the relevance of this mechanism to our experimental system.

3.2 EXPERIMENTAL PROCEDURES

3.2.1 Materials

Solvents were HPLC grade or LC-MS grade, as appropriate, and were purchased from EMD chemicals. Inorganic salts, HPLC buffers, and formamide were purchased from Sigma Aldrich, Fisher Scientific, or VWR, and were used as received. Bathophenanthroline, phenolphthalin, and most nucleobase standards were purchased from Sigma Aldrich; 8-aminoadenine was purchased from Tocris Bioscience and 2,6-diaminopurine from Carbosynth. Water was purified using a Barnstead nanoPure system and all HPLC buffers were filtered with 0.2 μ solvent filters (Millipore) prior to use. Pyrite was purchased from Alfa Aesar and ground into a powder prior to use. Industrial gases were purchased from AirGas South.

3.2.2 Reaction Conditions

Reactions were carried out in quartz test tubes in a custom-built isothermal aluminum heating block with temperature controlled to within ±0.5°C by a PID temperature controller. Reactions were stirred with Teflon-coated stir bars. For reactions conducted under an air atmosphere, test tubes were covered with glass to prevent evaporation. For reactions conducted under argon or carbon dioxide atmospheres, catalyst and formamide were added to the test tubes, then tubes were capped with rubber stoppers and reaction solutions were purged with the appropriate gas prior to the start of the reaction. Gas-containing balloons were also maintained over the test tubes during the reaction period. Two of the test tube wells in the heating block were slotted to allow

exposure to a low pressure mercury lamp (Pen-Ray) with primary emission at 254 nm and no emission of photons with wavelengths below 200 nm. All reaction solutions were prepared immediately prior to placement in the thermally-equilibrated reaction center.

Most reactions were carried out for 48 hours at 130°C with 200 mg of catalyst in 4 mL of formamide. For some reactions, catalyst was mixed with formamide at room temperature and vortexed, then centrifuged to pellet particulate matter. The supernatants of these solutions were then drawn off and reacted.

3.2.3 Sample Preparation

For HPLC and LC-MS/MS analysis, 200 μL aliquots of crude reaction mixtures were dried under vacuum at 60°C prior to injection, as formamide interfered significantly with chromatography.

3.2.3.1 LC-MS/MS

For LC-MS/MS, dried aliquots of formamide reaction mixtures were resuspended in 200 μL water, agitated with a vortex mixer, and centrifuged to pellet insoluble products. 5 μL aliquots of the supernatants of these centrifuged samples were added to 285 μL of acetonitrile and 10 μL of a solution containing the internal standard 2-aminopurine. For quantitation purposes, stock solutions of authentic standards of a variety of purines, pyrimidines, and triazines were also prepared. At least five different concentrations of each, in the same 95:5 acetonitrile:water mixture as for reaction samples, were injected into the LC for the purpose of establishing calibration curves. Spiking of authentic standards into reaction mixtures was also performed to confirm product identifications.

3.2.3.2 HPLC

For HPLC, dried aliquots of formamide reaction mixtures were resuspended in 200 μL of 250 mM phosphate buffer, diluted with water, agitated with a vortex mixer, and centrifuged in 0.2 μ cut-off spin filters (Pall Biosciences) to remove insoluble products and iron phosphate precipitates. In some cases, samples were spiked with authentic standards prior to injection.

3.2.3.3 Spectrophotometric Bathophenanthroline Assay

To determine the concentration of dissolved ferric and ferrous iron, a modified version of the bathophenanthroline assay was used [223, 224]. Bathophenanthroline (Figure 3.4) will chelate ferrous iron but not ferric iron; the Fe^{2+}·bathophenanthroline complex is red in color and strongly absorbs light at 533 nm.

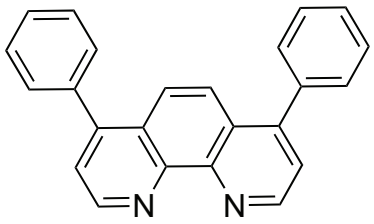

Figure 3.4 Bathophenanthroline (4,7-diphenyl-1,10-phenanthroline).

All glassware was soaked overnight in 50% aqueous HCl (v/v) to remove trace iron. 33.4 mg bathophenanthroline was dissolved in 100 mL of a 1:1 mixture of water and ethanol. 15 g of L-ascorbic acid was dissolved in 100 mL of water to which 0.1 mL of concentrated HCl was added. 4 mL of the prepared bathophenanthroline solution was added and the mixture was shaken. After 30 minutes, 10 mL of isoamyl alcohol was added and the solution was left at room temperature overnight. The upper organic layer,

which was pink in color, was then discarded. 100 mL of a 10% (w/v) aqueous solution of sodium acetate was prepared and 0.1 mL of concentrated HCl was added. 4 mL of the prepared ascorbic acid solution was added and the mixture was shaken. 4 mL of the bathophenanthroline solution was then added and the solution was again mixed. After 30 minutes, 10 mL of isoamyl alcohol was added and the solution was left at room temperature overnight. The upper organic layer was discarded. Reagent solutions were kept for no longer than a week.

An iron (II) chloride stock solution was prepared by dissolving 226.9 g $FeCl_2$ in iron-free water in a 1 L volumetric flask, adding 2.5 mL of concentrated sulfuric acid, and filling to the mark with iron-free water. A calibration curve was constructed using different amounts of this iron (II) chloride stock diluted to a final volume of 35 mL. 1.4 mL of ascorbic acid solution and 35 µL of concentrated HCl were added to each sample. Solutions were heated in a steam bath for 2 hours and cooled to room temperature. 1.75 mL of the sodium acetate solution and 1.4 mL of the bathophenanthroline solution were added, followed by thorough mixing. After 30 minutes, 5.5 mL of isoamyl alcohol were added and the samples were developed overnight at room temperature. A UV-Vis absorbance spectrum was then recorded for each concentration of iron and a simple regression analysis performed to calculate the equation of the calibration curve.

Aliquots of formamide reactions were removed at various time points during the formamide reactions and assayed using bathophenanthroline to detect total iron. These aliquots – usually only a few microliters were required – were diluted to 35 mL and then the procedure for treatment of iron samples used for calibration curve construction was followed to calculate total iron (ferrous and ferric). Identical aliquots were subjected to

all of these steps except for the ascorbic acid reduction and steam bath treatment to determine iron (II) concentrations. Iron (III) concentrations were then calculated by subtraction. This procedure is reported to result in 82% reduction of iron (III) to iron (II) [223] and values reported for ferric iron should therefore be considered as lower limits.

3.2.3.4 Spectrophotometric Phenolphthalin Assay

To determine the presence of cyanogen in reaction samples, a spectrophotometric assay in which phenolphthalin is transformed to phenolphthalein (Figure 3.5) was employed; phenolphthalein is a commonly-used indicator which takes on a pink color in basic solutions.

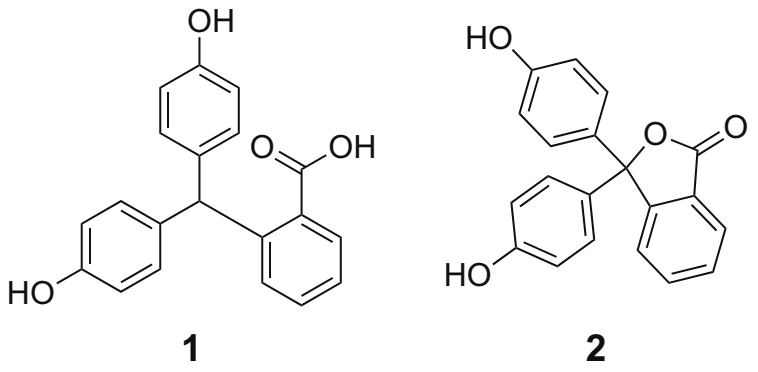

Figure 3.5 Phenolphthalin (**1**) and phenolphthalein (**2**). In the presence of copper (II) and cyanide, colorless phenolphthalin is converted to phenolphthalein, which is pink in basic solutions.

Cyanogen, $(CN)_2$, can be produced from cyanide ion in the presence of the copper (II) ion (Table 3.2). The phenolphthalin assay is carried out in basic solution to prevent formation of HCN gas and to ensure the proper pH for phenolphthalein color development; copper (II) hydroxide, however, is insoluble. EDTA in the reaction solution will complex the copper ions and prevent the precipitation of $Cu(OH)_2$ [218].

Table 3.2 Reactions of copper (II) and phenolphthalin. This reaction takes place in basic solution and EDTA is added to complex copper and prevent precipitation of copper (II) hydroxide. Adapted from reference [218].

Detection of Cyanide in Aqueous Solutions
$Cu^{2+} + 4CN^- \rightleftharpoons [Cu(CN)_4]^{2-}$
$2[Cu(CN)_4]^{2-} + 2CN^- \rightleftharpoons 2[Cu(CN)_4]^{3-} + (CN)_2$
$(CN)_2 + \text{phenolphthalin} \rightleftharpoons 2CN^- + \text{phenolphthalein} + 2H_3O^+$

Reagent solutions were prepared as follows: 19.068 g of sodium tetraborate decahydrate was dissolved in water and titrated to pH 10.8 with a 0.5 M NaOH solution; this was diluted to 2 L and used as the buffer for all phenolphthalin reaction solutions. 1.777 g phenolphthalin was dissolved in 500 mL of 50% ethanol in water (v/v). 4.187 g of $Cu(NO_3)_2 \cdot 2.5H_2O$ was dissolved in 500 mL of water; 3.359 g of disodium EDTA dihydrate was added to this solution.

The presence of cyanide in formamide reaction solutions was detected by diluting 100 μL of formamide solution with 1 mL of borate buffer in a microcentrifuge tube. 45 μL of phenolphthalin solution was added and the microcentrifuge was agitated with a vortex mixer. After 3 minutes, 25 μL of the copper/EDTA solution was added, the solution agitated again, and a UV-Vis spectrum immediately recorded.

Accurate quantitation of cyanide was not possible using this method because absorbance tails of some diluted reaction samples extended well past 533 nm where phenolphthalein has an absorbance maximum; this assay provided qualitative results only.

3.2.3.5 NMR

Raw formamide reaction mixtures were removed from heat, allowed to cool to room temperature, and were directly loaded into NMR tubes with no further processing or sample preparation.

3.2.4 Instrumental Methods

3.2.4.1 LC-MS/MS

3.2.4.1.1 Liquid Chromatography

LC separations of formamide reaction mixtures were achieved on an Agilent 1100 binary HPLC system. A Merck SeQuant ZIC-HILIC column (PEEK, 100 × 2.1 mm, 5 μm particle size, 200 Å porosity) was used for separations. The column was maintained at 30°C with a flow rate of 0.1 mL/min. The injection volume was 30 μL and each sample was injected twice, with different MRM transitions monitored in each injection.

Solvent A was acetonitrile; solvent B was 2.5 mM ammonium formate and 25 mM formic acid in water. The following gradient was used: 0-20 minutes, 5-20% B; 20-35 minutes, 20% B; 35-36 minutes, 20-5% B; 36-45 minutes, 5% B.

3.2.4.1.2 Multiple Reaction Monitoring

MRM was performed on a Micromass Quattro LC triple quadrupole mass spectrometer (see Table 3.3 for optimal MRM parameters). The source block temperature was 100°C and the desolvation temperature was 150°C. The dwell time for each MRM transition was 0.2 seconds and the interchannel delay was 1.8 seconds. Optimal parameters for detection of nucleobase standards (Figure 3.6) were determined by direct injection of standards into the mass spectrometer; these are given in Table 3.3. Positive ion mode was employed throughout.

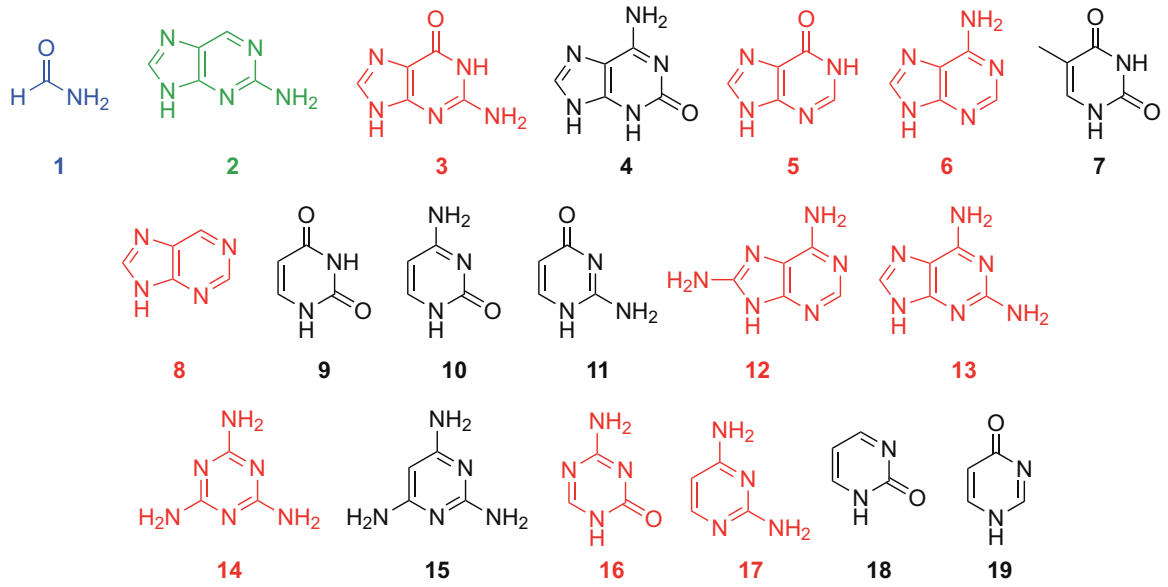

Figure 3.6 Structures of LC-MS/MS standards. Compounds in red were detected in formamide reactions conducted in the presence of pyrite mineral; compounds in black were not detected. **1** formamide, **2** 2-aminopurine (internal standard), **3** guanine, **4** isoguanine, **5** hypoxanthine, **6** adenine, **7** thymine, **8** purine, **9** uracil, **10** cytosine, **11** isocytosine, **12** 8-aminoadenine (6,8-diaminopurine), **13** 2,6-diaminopurine, **14** melamine, **15** 2,4,6-triaminopyrimidine, **16** 5-azacytosine, **17** 2,4-diaminopyrimidine, **18** 2-pyrimidinone, **19** 4(6)-hydroxypyrimidine.

Table 3.3 LC-MS/MS parameters used for optimal detection of nucleobase standards. Each sample was injected twice, with different mass transitions monitored in Methods A and B. Both A and B included monitoring of the 2-aminopurine internal standard (same precursor and product ion as adenine).

Standard	Precursor Ion (Da)	Product Ion (Da)	Cone Voltage (V)	Collision Voltage (V)	Retention Time (min)	Method
Guanine	151.8	134.8	40	25	23.5	A
Isoguanine	151.8	134.8	40	25	32.3	A
Hypoxanthine	136.9	109.8	50	30	15.6	A
Adenine	135.9	118.7	50	35	14.3	A, B
Thymine	126.9	109.9	45	20	6.0	A
Purine	120.9	93.9	40	20	8.6	A
Uracil	112.9	95.7	45	20	7.8	A
Cytosine	111.9	94.7	35	25	20.3	A
Isocytosine	111.9	94.7	35	25	16.7	A
8-Aminoadenine	150.7	133.8	45	30	24.8	B
2,6-Diaminopurine	150.7	133.8	45	30	18.0	B
Melamine	126.6	84.8	40	25	18.2	B
2,4,6-Triaminopyrimidine	125.7	67.0	35	50	26.6	B
5-Azacytosine	112.8	42.9	35	40	17.5	B
2,4-Diaminopyrimidine	110.6	93.9	40	30	24.8	B
2-Pyrimidinone	96.7	42.0	45	40	10.9	B
4(6)-Hydroxypyrimidine	96.7	42.0	45	40	7.6	B

Because chromatographic resolution is lost when more than eight precursor ion to product ion transitions are monitored in a single MRM run, the nucleobase standards were split into two groups. Each sample was injected twice, with seven transitions monitored in the first injection (Method A) and six transitions monitored in the second injection (Method B). The 135.9 Da to 118.7 Da transition was monitored in both methods; this transition represents adenine as well as the internal standard 2-aminopurine and allowed for peak height calibration. More than thirteen standards could be monitored because some compounds share transitions but have different retention times. Cytosine/isocytosine, guanine/isoguanine, 2,6-diaminopurine/8-aminoadenine, and 2-pyrimidinone/4(6)-hydroxypyrimidine make up four such pairs.

3.2.4.2 HPLC

HPLC analysis was performed on an Agilent 1100 HPLC system equipped with a DAD. Complete UV spectra from 190 to 400 nm were obtained for each time point in the chromatographic separation, though 260 nm proved to be the most convenient wavelength to monitor. Separations were reverse-phase and accomplished using a Phenomenex Synergi Polar RP column (250 × 4.6 mm, 4 μm particle size). The column was maintained at 30°C with a flow rate of 0.75 mL/min.

Solvent A was 20 mM sodium phosphate buffer at pH 7; solvent B was 10% acetonitrile in methanol (v/v). The following gradient was employed to achieve separation: 0-12 minutes, 5-10% B; 12-16 minutes, 10-20% B; 16-18 minutes, 20-40% B; 18-19 minutes, 40-5% B; 19-30 minutes, 5% B.

3.2.4.3 UV-Vis Spectroscopy

3.2.4.3.1 Bathophenanthroline Assay

UV-Vis spectra for the bathophenanthroline assay were collected in a precision quartz cell with a 1 cm path length (Hellma) and recorded on a Jasco V515 spectrophotometer. A sample spectrum is provided in Figure 3.7.

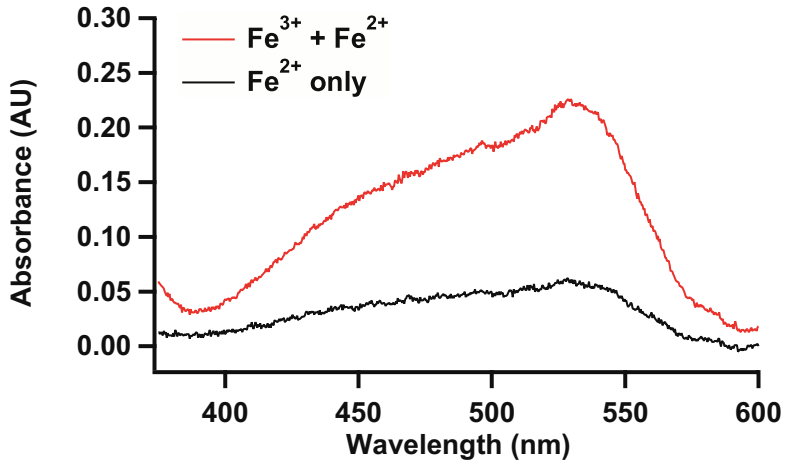

Figure 3.7 Sample spectra from the bathophenanthroline iron determination assay. Of two identical samples, one is reduced with ascorbic acid and the other is not. The unreduced sample provides an estimate of iron (II) concentration in solution, while the reduced sample provides an estimate of total iron (II) and iron (III). A value for iron (III) can then be obtained by subtraction.

3.2.4.3.2 Phenolphthalin Assay

UV-Vis spectra for the phenolphthalin assay were collected in a precision quartz microvolume cell with a 1 cm path length (Starna) and recorded on an Agilent 8453 UV-Vis spectrophotometer. A sample spectrum is provided in Figure 3.8.

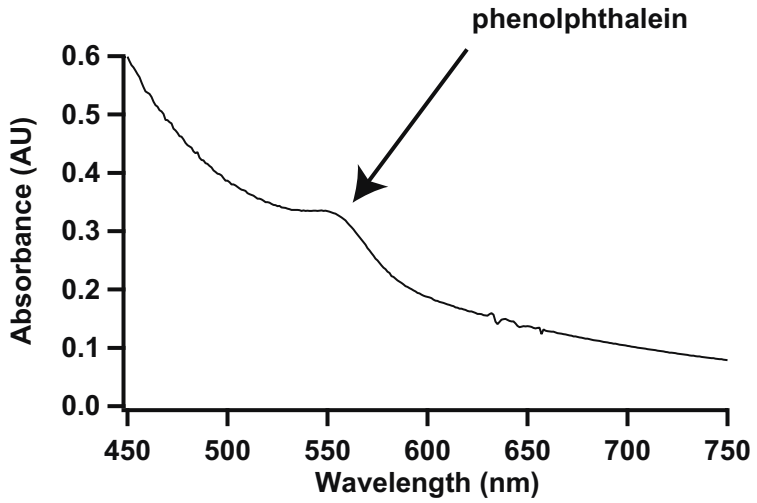

Figure 3.8 Sample spectrum from the phenolphthalin cyanide determination assay. The presence of an absorbance band with a maximum at 553 nm indicates conversion of phenolphthalin to phenolphthalein due to the presence of cyanide. Note that the sample matrix has a long absorbance tail and interferes with cyanide quantification.

3.2.4.4 NMR

^{1}H NMR spectroscopy was performed on a 400 MHz Varian Mercury Vx instrument. Spectra were acquired unlocked as single transients without a reference standard. A sample spectrum is provided in Figure 3.9.

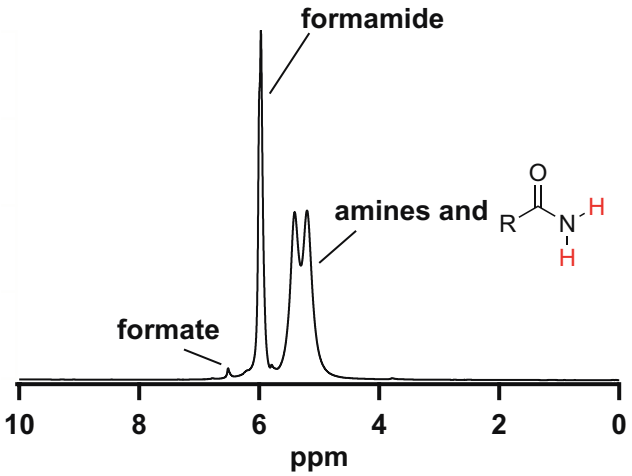

Figure 3.9 Sample ^{1}H NMR spectrum indicating the positions of formate, formamide, and amine/amide proton resonances. Relevant amide protons are shown in red. Spectrum was acquired as a single, unlocked transient. No reference standard was used.

3.3 RESULTS

3.3.1 Product Characterization

Of the 17 standards tested, nine products were identified under various reaction conditions; these include purine, adenine, hypoxanthine, guanine, 2,6-diaminopurine, 6,8-diaminopurine, 2,4-diaminopyrimidine, melamine, and 5-azacytosine. The yields and diversities of products generated were shown to depend on atmosphere, presence of catalyst, and UV irradiation.

3.3.1.1 Air Atmosphere

Thermal-only formamide reactions generated few products other than purine and the triazines 5-azacytosine and melamine. While compounds such as 5-azacytosine have the same pattern of exocyclic substituents as biomolecules such as cytosine, in general, they appear to be poorly-suited as recognition elements in nucleic acid helices [225].

In an air atmosphere, thermal and UV-irradiated formamide solutions exhibit similar product profiles, including the presence of adenine, guanine, and hypoxanthine

nucleobases – products typically not detected under thermal-only conditions. The UV-irradiated reactions with no catalyst and with pyrite mineral present generate nearly identical suites of products (Figures 3.10 and 3.11): adenine, 5-azacytosine, guanine, hypoxanthine, and melamine, in addition to purine.

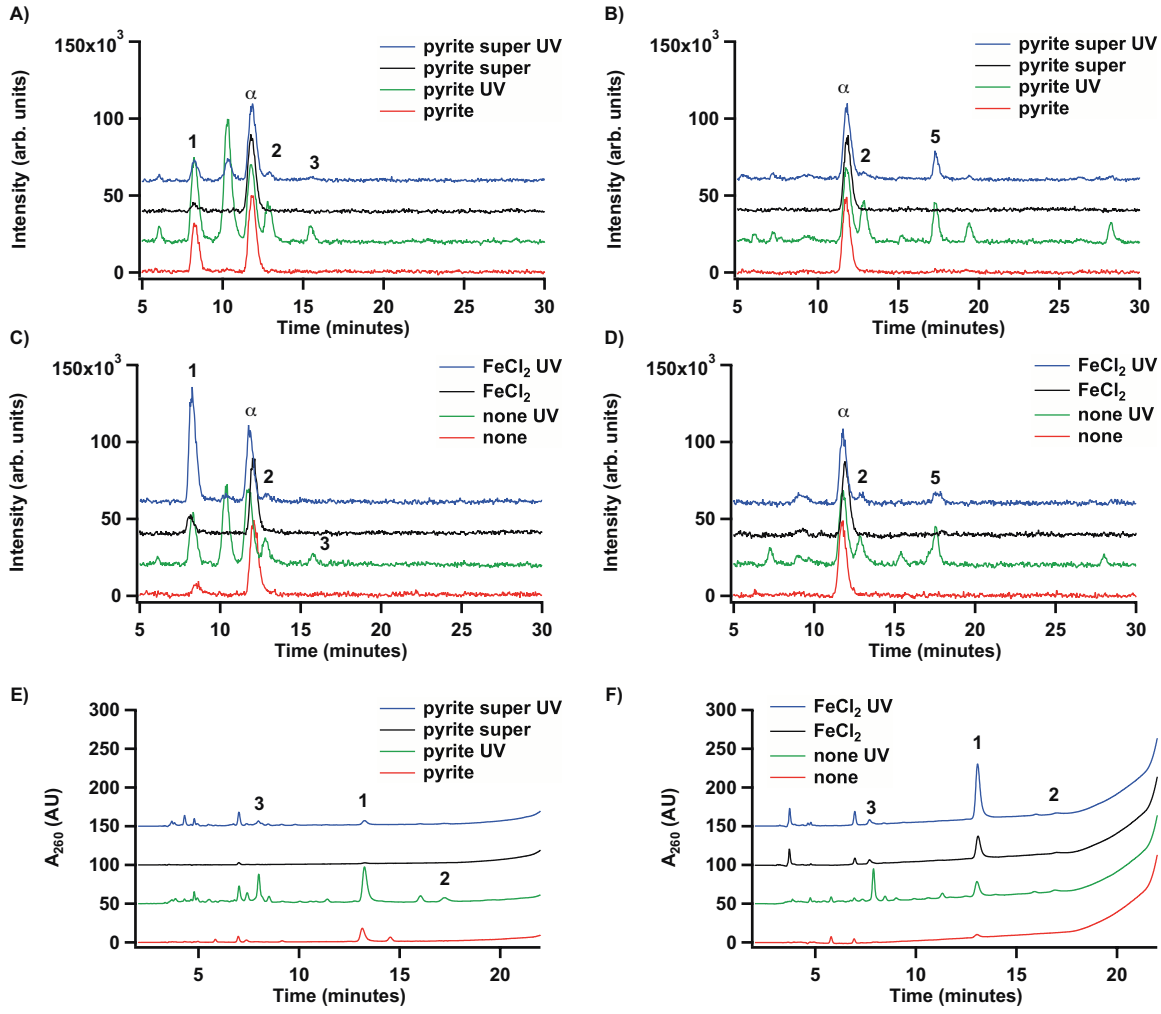

Figure 3.10 (A-D) LC-MS/MS and (E-F) HPLC chromatograms of formamide reactions conducted under an air atmosphere. Selected peaks are labeled: **1** purine, **2** adenine, **3** hypoxanthine, **5** 5-azacytosine, **α** 2-aminopurine (internal standard). (A) and (B) represent two injections of the same sample but monitoring of different MS/MS transitions; this is also true of (C) and (D). Note that (A-D) represent combined selected ion chromatograms and minor products may not be visible here, but can be seen in individual ion channels. See text and color-coded keys on individual panels for experimental details.

A pyrite "supernatant" sample in which pyrite is mixed with formamide and centrifuged, with only the supernatant of this mixture being heated and irradiated, displays reduced adenine and hypoxanthine yields and no melamine, while a reaction with abundant iron (II) ions in solution ($FeCl_2$ catalyst) generates no guanine or hypoxanthine and a comparable amount of melamine as a reaction with no catalyst or with solid pyrite present (Figure 3.11).

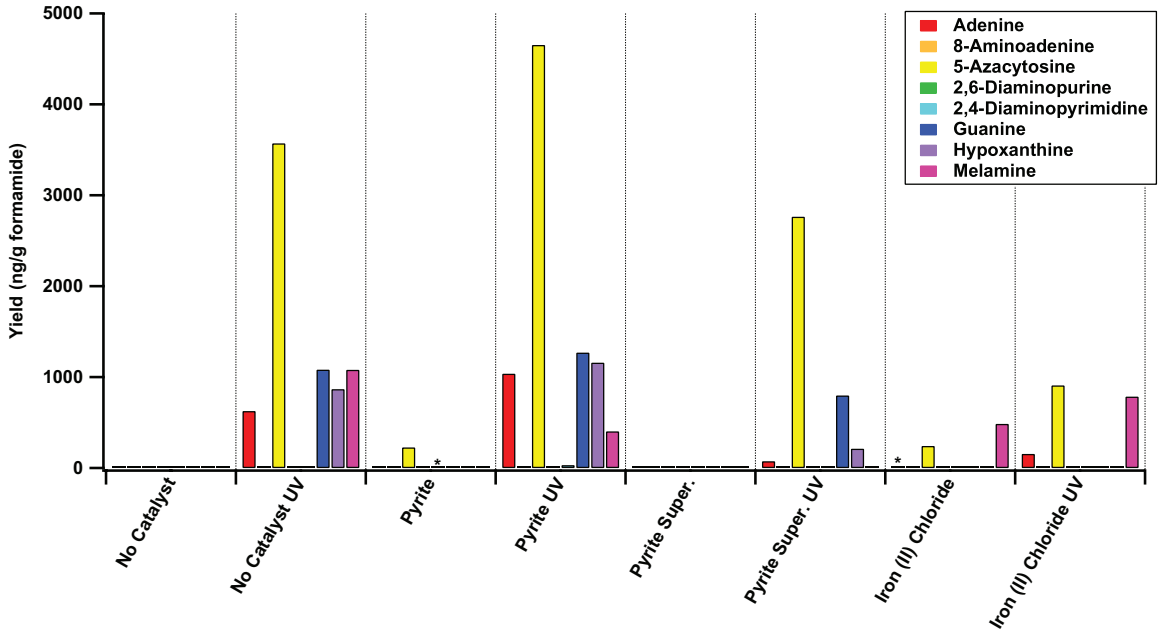

Figure 3.11 Yields of nucleobases and related compounds under an air atmosphere. In each case, 4 mL of formamide was heated at 130°C for 48 hours in the presence of 200 mg catalyst, when catalyst was present. For supernatant samples ("super." in figure legend), catalyst was mixed with formamide at room temperature, the solution was centrifuged, and the supernatant drawn off and reacted. This represents a case where a concentration of ions less than saturation are in solution and no solids are present. Some samples were irradiated with 254 nm UV light. Asterisks (*) indicate tentative identifications.

These results indicate that iron (II) in solution is not solely responsible for the products generated in formamide reactions, and in fact may contribute to the destruction of certain compounds. Further, we find that under many conditions, UV irradiation still has an equalizing effect, which supports our previous results [40].

3.3.1.2 Argon Atmosphere

In an effort to determine the contributions of atmospheric gases, we conducted formamide reactions after removing air, and hence oxygen, from the system. We chose an argon atmosphere as a control since argon is an inert gas. For thermal-only reactions, few products other than purine were generated, although an argon atmosphere does allow for the production of small amounts of adenine when pyrite mineral or $FeCl_2$ is present as a catalyst (Figure 3.12 and 3.13).

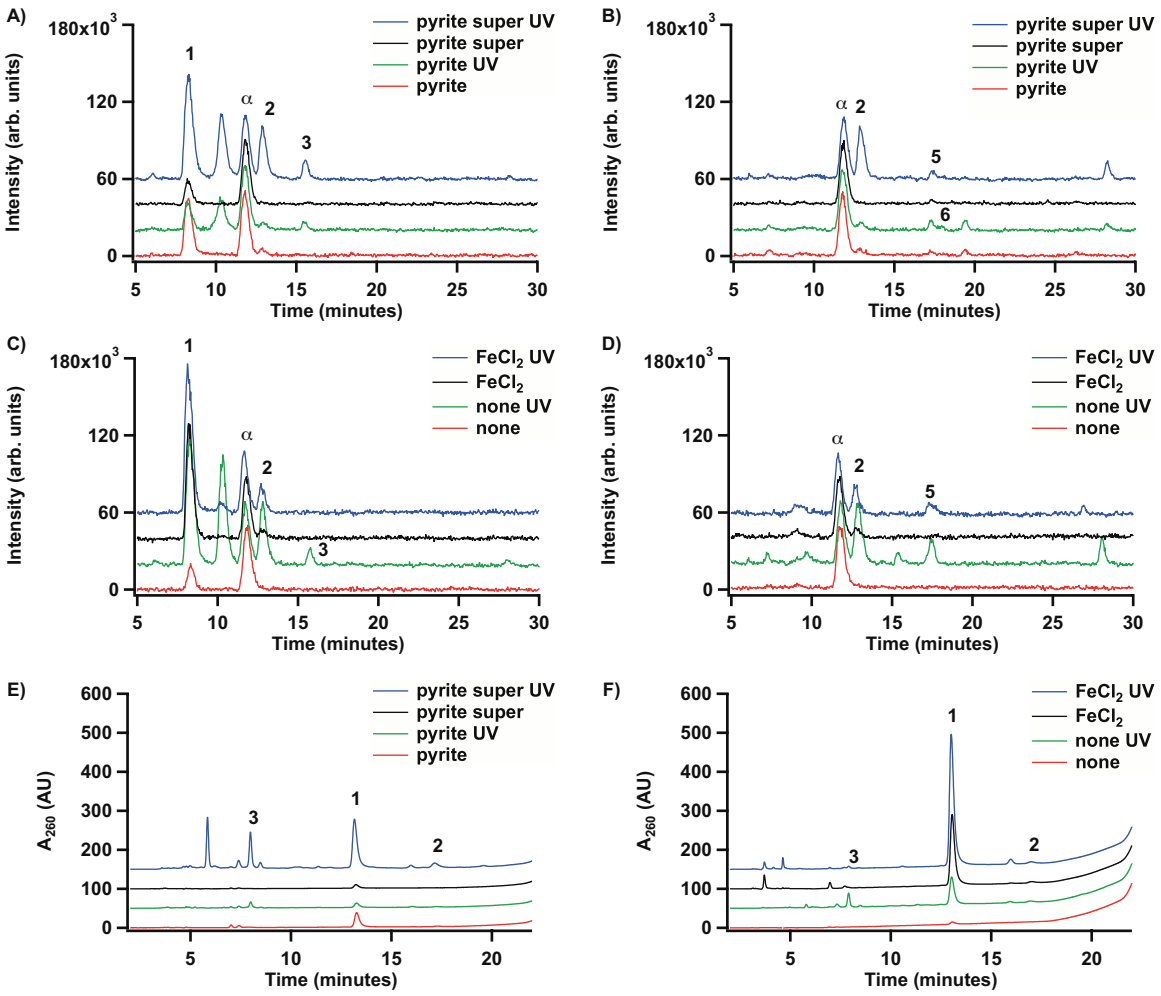

Figure 3.12 (A-D) LC-MS/MS and (E-F) HPLC chromatograms of formamide reactions conducted under an argon atmosphere. Selected peaks are labeled: **1** purine, **2** adenine, **3** hypoxanthine, **5** 5-azacytosine, **6** melamine, **α** 2-aminopurine (internal standard). (A) and (B) represent two injections of the same sample but monitoring of different MS/MS transitions; this is also true of (C) and (D). Note that (A-D) represent combined selected ion chromatograms and minor products may not be visible here, but can be seen in individual ion channels. See text and color-coded keys on individual panels for experimental details.

When samples are irradiated with 254 nm UV light under an argon atmosphere, however, the yields and diversities of products change dramatically. No guanine was detected under any condition tested under an argon atmosphere, though hypoxanthine is still generated in appreciable amounts in most cases. Further, 5-azacytosine is no longer the dominant product (excluding purine) for the no catalyst and pyrite supernatant samples; yields of adenine and hypoxanthine surpass it (Figure 3.13). Additionally, 8-aminoadenine (6,8-diaminopurine) is detected for the first time under an argon atmosphere in the UV-irradiated no catalyst reaction, as is a trace amount of 2,6-diaminopurine (Figure 3.13).

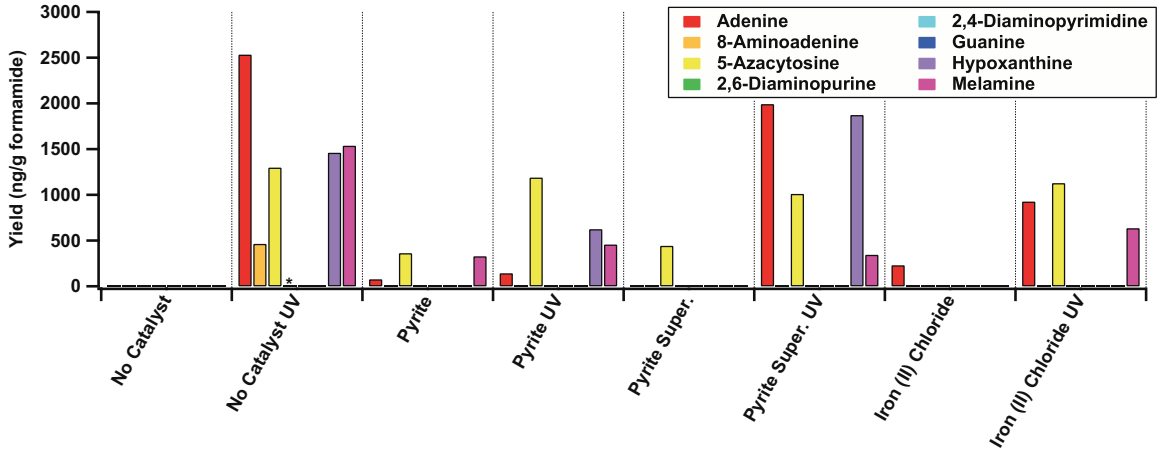

Figure 3.13 Yields of nucleobases and related compounds under an argon atmosphere. In each case, 4 mL of formamide was heated at 130°C for 48 hours in the presence of 200 mg catalyst, when catalyst was present. For supernatant samples ("super." in figure legend), catalyst was mixed with formamide at room temperature, the solution was centrifuged, and the supernatant drawn off and reacted. This represents a case where a concentration of ions less than saturation are in solution and no solids are present. Some samples were irradiated with 254 nm UV light. Asterisks (*) indicate tentative identifications.

With the exception of the increased amount of hypoxanthine generated in several cases, products with exocyclic amino groups appear to be preferred under an argon atmosphere to those with keto or hydroxyl substituents. It should be remembered, however, that hypoxanthine may be a hydrolysis product of adenine. These results strongly suggest that the synthesis of oxygen-containing products from UV-irradiated formamide solutions is enhanced in oxidizing or oxygen-containing atmospheres.

3.3.1.3 Carbon Dioxide Atmosphere

Experimental and observational evidence from the literature leads to varied, sometimes conflicting, conclusions about the composition of the atmosphere of a prebiotic Earth. According to several current hypotheses, carbon dioxide was likely to have been present in Earth's atmosphere at the time life originated [58, 59]. While it was not possible with our experimental apparatus to conduct reactions under precise mixtures of gases, we conducted formamide experiments in a pure carbon dioxide atmosphere as a model for a prebiotic atmosphere with a large partial pressure of CO_2.

For thermal-only formamide reactions conducted in the presence of CO_2, few products are generated in the pyrite mineral sample, in the pyrite supernatant sample, or in the no catalyst control (Figures 3.14 and 3.15). An iron (II) chloride reaction, meanwhile, is biased in favor of the triazines 5-azacytosine and melamine, with small amounts of adenine and 2,6-diaminopurine being produced, as well (Figure 3.15).

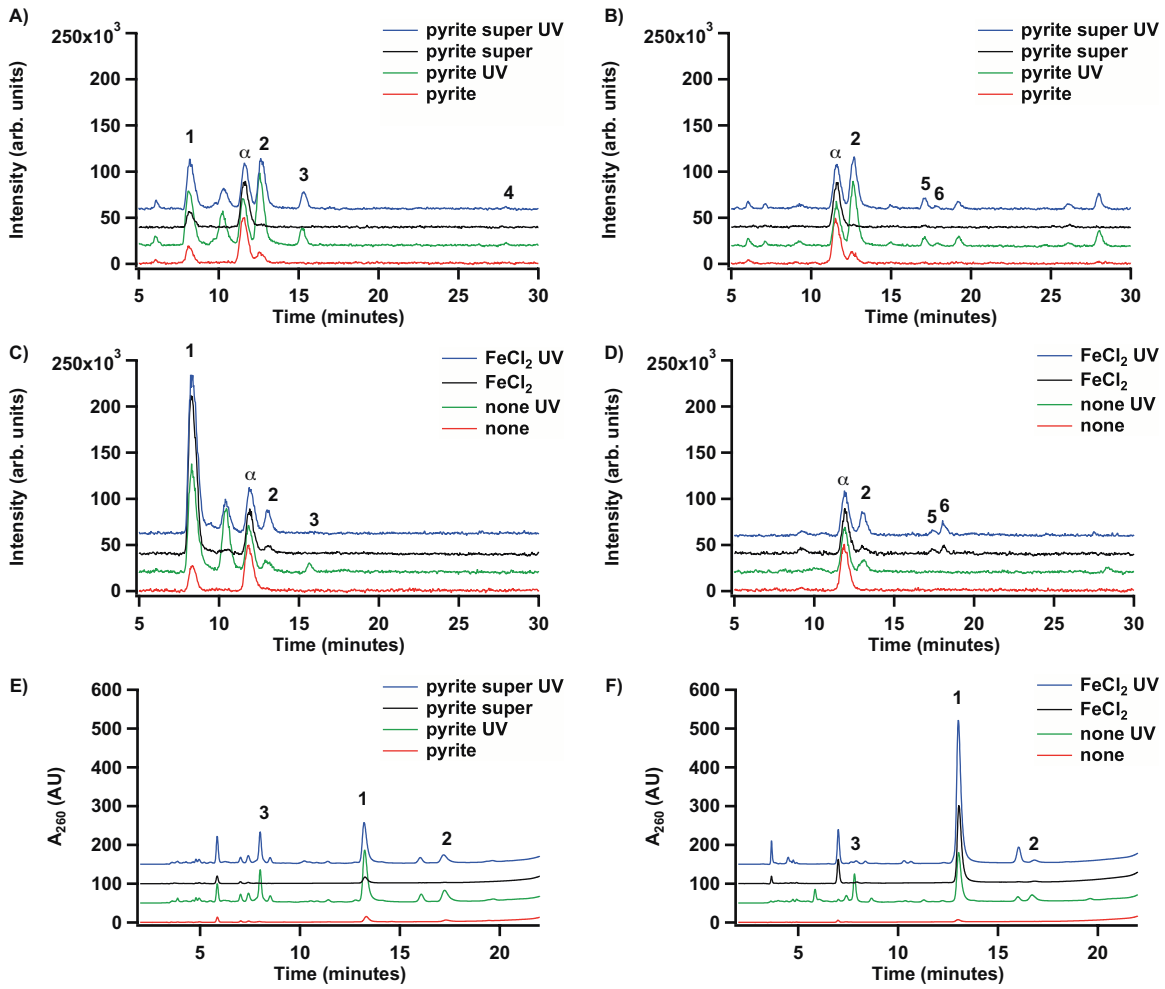

Figure 3.14 (A-D) LC-MS/MS and (E-F) HPLC chromatograms of formamide reactions conducted under a carbon dioxide atmosphere. Selected peaks are labeled: **1** purine, **2** adenine, **3** hypoxanthine, **4** guanine, **5** 5-azacytosine, **6** melamine, **α** 2-aminopurine (internal standard). (A) and (B) represent two injections of the same sample but monitoring of different MS/MS transitions; this is also true of (C) and (D). Note that (A-D) represent combined selected ion chromatograms and minor products may not be visible here, but can be seen in individual ion channels. See text and color-coded keys on individual panels for experimental details.

When UV irradiation is introduced to the system, however, the synergetic effects of catalyst, radiation, and atmosphere become apparent. While the control containing no catalyst generates moderate amounts of adenine and hypoxanthine, the pyrite and pyrite supernatant reactions generate the highest levels of adenine and hypoxanthine of all of

the reactions we performed; yields of both of these products exceed the yields of 5-azacytosine, which was not true under an air atmosphere (Figure 3.15). Further, guanine production is revived under a carbon dioxide atmosphere, and small amounts of 8-aminoadenine and 2,4-diaminopyrimidine are seen in both the pyrite and pyrite supernatant samples (Figure 3.15); these conditions result in the greatest diversity of products, including biologically-relevant products, of any reaction condition we explored.

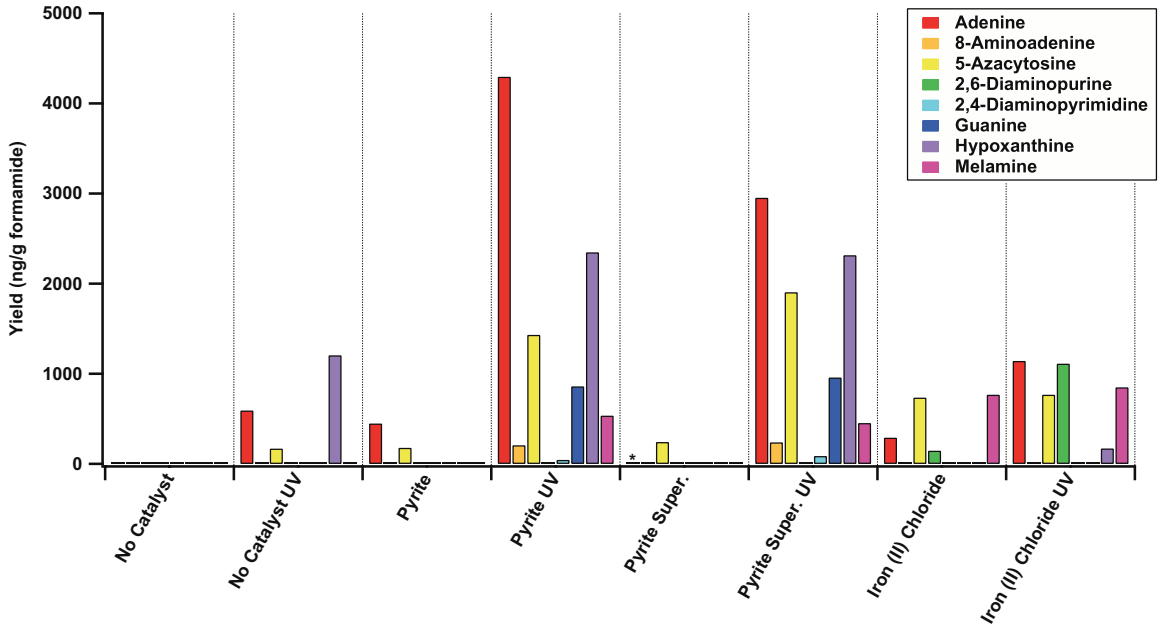

Figure 3.15 Yields of nucleobases and related compounds under a carbon dioxide atmosphere. In each case, 4 mL of formamide was heated at 130°C for 48 hours in the presence of 200 mg catalyst, when catalyst was present. For supernatant samples ("super." in figure legend), catalyst was mixed with formamide at room temperature, the solution was centrifuged, and the supernatant drawn off and reacted. This represents a case where a concentration of ions less than saturation are in solution and no solids are present. Some samples were irradiated with 254 nm UV light. Asterisks (*) indicate tentative identifications.

Further, these yields are not due to dissolved Fe^{2+} ions, as an $FeCl_2$ control reaction displays a different product diversity under the same conditions, showing no

guanine, low hypoxanthine, and a relatively high yield of 2,6-diaminopurine (Figure 3.15).

3.3.2 Competing Pathways of Purine and Adenine Formation

The unified mechanism of purine and adenine production from formamide [217] could be active in our reaction mixtures; we examined yields of unsubstituted purine, conversion of formamide to formate, and the presence of cyanide in our reactions to gain a greater understanding of the role of this particular mechanism in producing the results we have observed.

3.3.2.1 Purine Yields

Unsubstituted purine is a characteristic product of formamide chemistry [36, 40, 80]. Yields of reduced purine, especially in relation to adenine yields, can provide valuable information on the reaction mechanisms active under various experimental conditions. In particular, increased formation of adenine coupled with reduced formation of purine suggests increased concentrations of the cyanide ion in solution, whereas an adenine:purine ratio biased in favor of purine indicates the presence of a substantial amount of hydride-generating formate ion [217].

Purine yields for all reactions are generally several orders of magnitude higher than adenine yields (Figure 3.16). However, adenine yields, relative to purine yields, are the highest for the pyrite and pyrite supernatant reaction samples which are UV-irradiated under a carbon dioxide atmosphere.

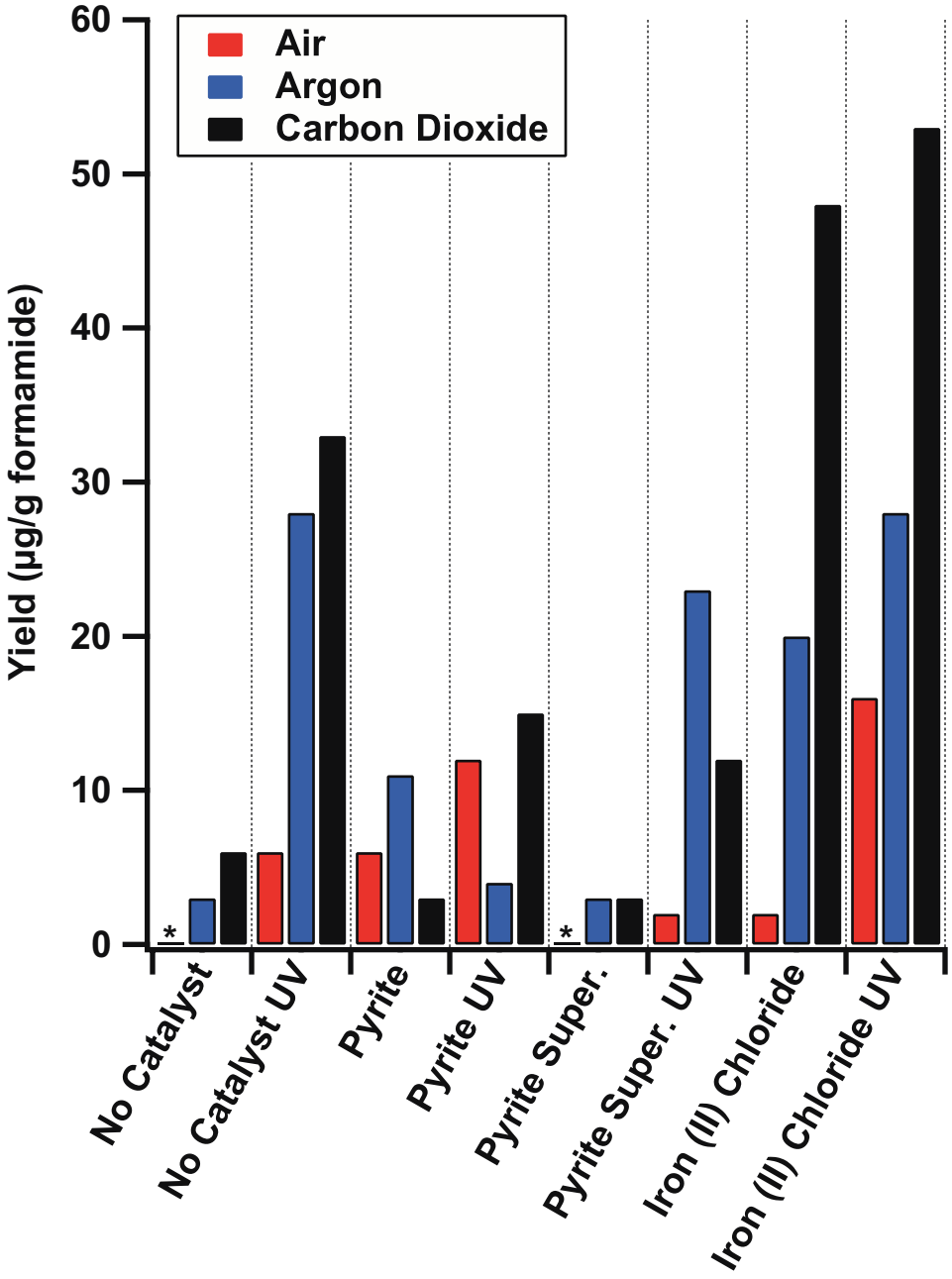

Figure 3.16 Purine yields under various reaction conditions. In each case, 4 mL of formamide was heated at 130°C for 48 hours in the presence of 200 mg catalyst, when catalyst was present. For supernatant samples ("super." In figure legend), catalyst was mixed with formamide at room temperature, the solution was centrifuged, and the supernatant drawn off and reacted. This represents a case where a concentration of ions less than saturation are in solution and no solids are present. Some samples were irradiated with 254 nm UV light. Asterisks (*) indicate tentative identifications.

3.3.2.2 Formate Quantification

Measurements of the conversion of formamide to formate reveal only small differences in formate production overall, with an air atmosphere leading to the highest and lowest percentages of conversion, and a carbon dioxide atmosphere leading to only small differences in conversion percentage over the range of reaction conditions (Table 3.4).

Table 3.4 Amount of formate present (as total percentage of bulk solvent) in formamide reaction solutions after a 48 hour reaction period. Formate could not be measured in $FeCl_2$ samples by NMR due to large amounts of dissolved iron.

	Atmosphere		
Reaction	Air	Argon	CO_2
No Catalyst	0.87	1.76	1.70
No Catalyst with UV	0.49	1.57	1.19
Pyrite	0.91	1.67	1.50
Pyrite with UV	1.07	1.57	1.43
Pyrite Super.	2.15	1.47	1.38
Pyrite Super. with UV	1.63	0.73	1.30

These results suggest that in a complex reaction matrix which includes a mineral component, UV irradiation, and controlled atmospheric gases, the role of the formate ion may not be as straightforward as in solutions that lack these components. Further, it should be noted that small amounts of DAMN in molten ammonium formate are known to produce adenine upon heating [116], so formate may contribute to adenine production, as well.

3.3.2.3 Cyanide Determination

We were unable to quantify absolutely levels of cyanide in our reaction solutions due to significant spectral interference from our formamide reaction sample matrices. Low levels of cyanide production were noted in all experimental samples; however, certain combinations of reaction conditions were shown to generate more cyanide than others (Table 3.5).

Table 3.5 Cyanide determination. (✓) indicates a cyanide peak was observed spectrophotometrically after three minutes (see Section 3.2.4.3.2). "Trace" indicates that the reaction solution turned pink, indicating the production of phenolphthalein, after an extended period of time (overnight to several weeks). See text for additional information.

	Atmosphere		
Reaction	Air	Argon	CO_2
None	✓	trace	trace
None with UV	✓	trace	trace
Pyrite	trace	✓	trace
Pyrite with UV	trace	✓	trace
Pyrite Super.	trace	trace	✓
Pyrite Super. with UV	trace	✓	trace
$FeCl_2$	trace	trace	trace
$FeCl_2$ with UV	trace	trace	trace

There are two phases of color development in the spectrophotometric phenolphthalin assay we used to detect cyanide. The initial fast production of pink color provides an accurate measure of cyanide present, while a longer-term emergence of color indicates the recycling of cyanide as cyanogen is reductively split in the reaction that converts phenolphthalin to phenolphthalein [218]. A reaction that produces even a trace

amount of cyanide – unmeasurable by this assay – will turn pink if left to react long enough.

There are few, if any, traits unifying the six reactions which produce measurable cyanide (represented by checkmarks in table 3.5). They occur under air, argon, and carbon dioxide atmospheres, and may or may not be UV irradiated; solid pyrite may or may not be present. It is interesting to note that while a variety of anions have been demonstrated not to interfere with this assay [218], solutions containing cations which may chelate cyanide were not tested for interference; further, only reaction solutions with significant amounts of dissolved iron (namely, $FeCl_2$-catalyzed reactions) displayed no measurable cyanide after the initial color development stage. This may be due to the sequestration of cyanide in iron complexes such as Prussian blue ($Fe_7(CN)_{18}$), leaving it unavailable to undergo reaction with phenolphthalin.

3.3.3 Solubility of Iron Compounds

Previous studies of formamide chemistry have evaluated the catalytic capabilities of solid minerals, but have not directly considered the contributions of solution-phase chemistry. Because micromolar amounts of soluble Fe^{2+} are likely to have been available on the prebiotic Earth [95], we conducted reactions with $FeCl_2$, an iron (II) salt which is soluble in formamide, to distinguish the effects of soluble iron from those of solid pyrite mineral. When solid pyrite is heated in formamide, little ferric or ferrous iron can be detected in solution as determined by the bathophenanthroline assay (see Sections 3.2.3.3 and 3.2.4.3.1), even after an extended period of time. Total dissolved iron, represented as the sum of ferric and ferrous iron in solution, for an $FeCl_2$-catalyzed reaction, however,

levels off at about 4 μg iron/μL formamide after 48 hours, with slightly more dissolved Fe^{2+} than Fe^{3+} at the end of the reaction period (Figure 3.17).

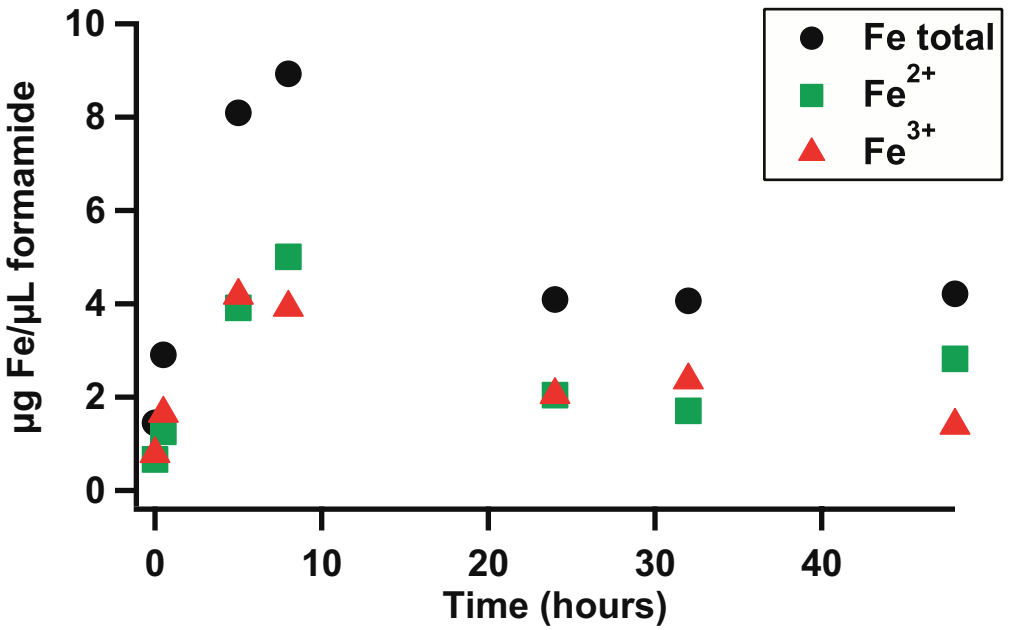

Figure 3.17 Ferric and ferrous iron in formamide solutions after heating at 130°C in the presence of $FeCl_2$ as determined by the bathophenanthroline spectrophotometric assay. No dissolved iron was detected in pyrite solutions.

While the effects of the disulfide anion have not been examined and cannot be discounted, it is clear from our LC-MS/MS analysis that pyrite's unique ability to catalyze nitrogen heterocycle formation from formamide is not due solely to the Fe^{2+} cation, and that the mineral surface or matrix also likely contributes.

3.4 DISCUSSION

3.4.1 Synergetic Effects of Pyrite, UV Irradiation, and Atmospheric Gases

This study has demonstrated the combined effects of a pyrite mineral catalyst, UV irradiation, heat, and atmosphere on the yield and diversity of products generated in formamide reactions; this work represents the first study of formamide chemistry not conducted in an air atmosphere or under high vacuum. Further, we have examined the role of soluble iron versus iron in a mineral matrix and shown that solution phase chemistry does not fully account for the catalytic behavior of solid pyrite. It should be noted that centrifugation of pyrite suspended in formamide might not remove microfine particles from supernatant samples. The catalytic abilities of nanoparticles are well-documented and a topic of much current research in the field of green chemistry [226]; nanoparticles in pyrite and pyrite supernatant samples may be important catalytic species. Hydrothermal vents are reported to distribute pyrite nanoparticles throughout the oceans even today [227]; it is reasonable to assume a similar process may have been acting since the time of the origin of life.

All of the vital reaction components – heat, UV irradiation, formamide, pyrite, and a carbon dioxide atmosphere – must be combined in a single reaction vessel to achieve maximum yield and diversity of products, including a preference for the biologically-relevant guanine, adenine, and hypoxanthine over 5-azacytosine. Purine yields are somewhat reduced under these combined conditions, as well.

3.4.2 Surface Versus Solution-Phase Catalysis

It is clear that chemistry occurring in pyrite and pyrite supernatant samples does not derive from dissolved iron (II) ions; not only can no iron (II) be detected in pyrite

reaction samples, but a different suite of products is generated when high amounts of soluble iron are present. $FeCl_2$-catalyzed reactions generate few products under air and argon atmospheres, and a different suite of products including high yields of melamine and significant production of 2,6-diaminopurine; these results do not resemble those for the pyrite systems. Further, purine yields are increased in $FeCl_2$-containing reactions under argon and CO_2 atmospheres, relative to pyrite reactions.

Pyrite may be most catalytically-active, however, as nanoparticles that resist removal from solution by centrifugation, as indicated by the high yields of a variety of products in pyrite supernatant samples.

3.4.3 Identification and Significance of New Products

Several of the products identified for the first time from formamide reactions as a result of this work, including 2,6-diaminopurine and 8-aminoadenine, have also been detected in meteorite samples; these are reportedly consistent with ammonium cyanide condensation chemistry [119]. Further, we have identified melamine for the first time in a formamide reaction. This triazine is a trimer of cyanamide (Figure 3.18), a molecule known to catalyze the condensation of amino acids into dipeptides [228, 229] and the phosphorylation of nucleosides [143]. Cyanamide has also been suggested as an intermediate in the prebiotic synthesis of activated pyrimidine nucleotides [144]. Carbodiimide, an isomer of cyanamide, has previously been reported as a product of formamide reactions [36].

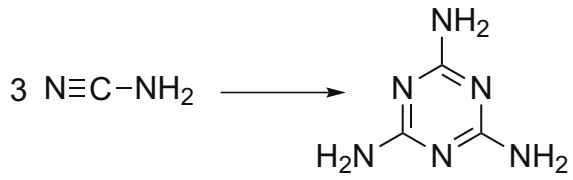

Figure 3.18 Three cyanamide molecules can combine to form melamine.

3.4.4 Mechanism of Purine and Adenine Synthesis

Our assays for conversion of formamide to formate and for the detection of cyanide in solution were able to demonstrate the presence of both of these important ions in our reaction mixtures. However, the presence or abundance of these species did not correlate to yields of purine or adenine. While the unified mechanism of adenine and purine synthesis is compelling [217], it may compete with other reactions in our system.

3.5 CONCLUDING REMARKS

While a synergy clearly exists among pyrite, heat, UV irradiation, and atmospheric composition with respect to the production of nucleobases from formamide, all of our experiments were conducted with 254 nm UV irradiation. Formamide has an absorbance tail that extends past 254 nm [40, 202] and is possibly the photoactive species in these reactions. It would be instructive to conduct a parallel set of experiments using longer-wavelength UV irradiation, where the pyrite absorbs but the formamide does not.

While little agreement exists about the composition of the Earth's early atmosphere [63, 106, 230-232], our results suggest that important biological molecules and their analogs can be synthesized from formamide under a variety of atmospheric conditions. We propose that finding experimental conditions with compatible chemistry for synthesizing life's molecules can be of great assistance in setting limits on the range of permissible atmospheres and geologies that may have contributed to life's origin.

CHAPTER 4

FURTHER INVESTIGATIONS OF PREBIOTIC FORMAMIDE CHEMISTRY

4.1 INTRODUCTION

Our interest in the abilities of iron salts to catalyze the production of nucleobases and nucleobase analogs from formamide solutions led us to consider the contributions of anions to formamide chemistry. Many iron-containing minerals also include phosphate groups; further, iron salts differ dramatically as to their solubilities in formamide. Additionally, performing formamide reactions with a variety of iron salts and minerals should provide further information as to the role of iron as a catalyst in its ferrous and ferric charge states.

Previously, we monitored the production of adenine and hypoxanthine when DAMN or the substituted imidazoles AICA and AICN were spiked into reaction mixtures. These molecules each possess four carbon atoms; it is also of great interest to consider the effects of smaller molecules on the production of nucleobases from formamide. It is unlikely that pools of neat formamide existed on early Earth; while evaporation of water from mixed water/formamide systems provides a mechanism for concentrating formamide (the "drying pool" scenario), it is highly likely that other one-, two-, and three-carbon molecules may have been present in these pools. These small molecules may also have participated in and have affected the outcomes of formamide reactions.

4.1.1 The Importance of Phosphate Minerals

4.1.1.1 Phosphates in Contemporary Biology

Phosphate groups are ubiquitous in modern biology; in addition to appearing as linkers in the backbones of RNA and DNA molecules, phosphates are seen in the phospholipids that make up our cell membranes, in energy carriers such as ATP and GTP, in photosynthetic and glycolytic intermediates, and in coenzymes (Figure 4.1). Additionally, phosphorylation of proteins is an important post-translational modification.

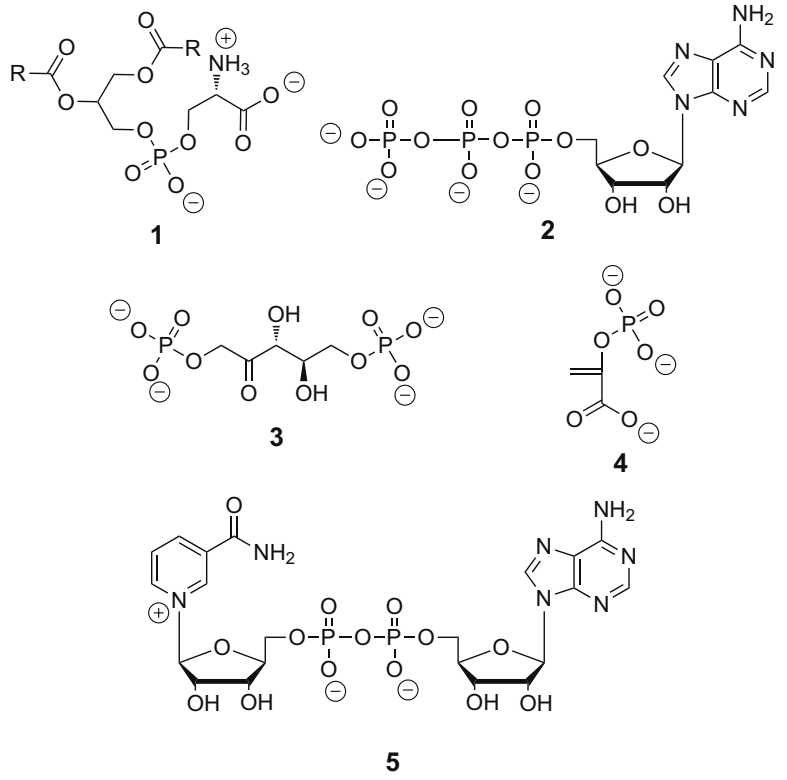

Figure 4.1 Examples of phosphate-containing molecules in contemporary biology. **1** phosphatidylserine, a phospholipid head group usually found on the cytosolic side of cell membranes (R may be one of several hydrophobic tails); **2** adenosine triphosphate (ATP), coenzyme energy carrier in cells; **3** ribulose-1,5-bisphosphate, important in the dark reactions of photosynthesis; **4** phosphoenolpyruvate (PEP), a glycolytic intermediate, **5** nicotinamide adenine dinucleotide + (NAD+), an electron carrier coenzyme involved in redox reactions in cells. Phosphate groups are also present in the backbones of DNA and RNA and many proteins are phosphorylated post-translationally [233].

The phosphoester bonds found in life are kinetically stable to hydrolysis for a period of up to 1000 years [234, 235]; phosphate groups may have been preceded as linkers by any number of molecules, including glyoxylate and/or glycolaldehyde [152]. Today, most inorganic phosphate is restricted to insoluble minerals; phosphate is thus often a limiting reagent in ecosystems [83, 236]. There exists great interest in the prebiotic chemistry community in determining how the first phosphate groups were incorporated into what would become the molecules of life.

4.1.1.2 Prebiotic Phosphate Sources

Numerous prebiotically plausible sources of phosphate have been proposed. Many nucleoside phosphorylation reactions employ simple potassium phosphates [83]; others have successfully employed minerals as phosphate donors [81]. Some transition metal phosphates are hypothesized to have been available on the primitive Earth [87]. Additionally, reduced forms of phosphate such as schreibersite ($(Fe,Ni)_3P$) may have been delivered exogenously via meteorites [87, 236]; these may have reacted with water to produce soluble, reactive phosphorus compounds [236].

4.1.1.3 Phosphate Minerals and Formamide Chemistry

Phosphate minerals have also attracted attention as catalysts for formamide chemistry [82]; a selection of phosphate mineral catalysts is provided in Table 4.1.

Table 4.1 Selected phosphate minerals and their elemental compositions.

Mineral	Elemental Composition
Augelite	$Al_2(PO_4)(OH)_3$
Wavellite	$Al_3(PO_4)_2(OH,F)_3 \cdot 5H_2O$
Hureaulite	$Mn_5(PO_3OH)_2(PO_4)_2 \cdot 4H_2O$
Vivianite	$Fe_3(PO_4)_2 \cdot 8H_2O$
Ludlamite	$(Fe,Mn,Mg)_3(PO_4)_2 \cdot 4H_2O$
Libethenite	$Cu_2(PO_4)(OH)$
Lazulite	$(Mg,Fe)Al_2(OH,PO_4)_2$
Vauxite	$FeAl_2(PO_4)_2(OH)_2 \cdot 6H_2O$
Childrenite	$(Fe,Mn)Al(PO_4)(OH)_2 \cdot H_2O$
Turquoise	$CuAl_6(PO_4)_4(OH)_8 \cdot 4H_2O$
Pyromorphite	$Pb_5(PO_4)_3Cl$

Upon heating formamide in the presence of various phosphate minerals, a rich and versatile reactivity is displayed; products of the phosphate mineral-catalyzed formamide reaction include purine and pyrimidine nucleobases, urea, a formylated amino acid, and carbodiimide, a known condensing agent [82] (Figure 4.2).

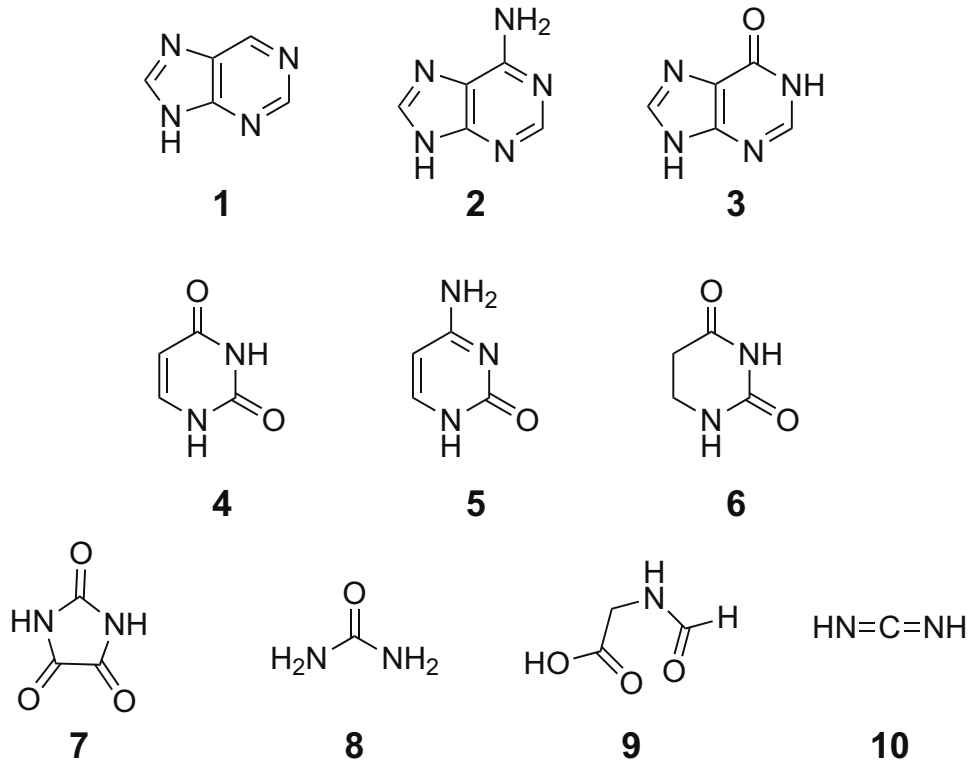

Figure 4.2 A selection of products generated when formamide is heated in the presence of phosphate minerals. **1** purine, **2** adenine, **3** hypoxanthine, **4** uracil, **5** cytosine, **6** 5,6-dihydrouracil, **7** parabanic acid, **8** urea, **9** N-formylglycine, **10** carbodiimide [82].

4.1.1.4 Iron Phosphates

Iron phosphates are of particular relevance because in addition to being possible sources of phosphate, their iron atoms offer the potential for catalysis of redox reactions. Vivianite is a photosensitive iron phosphate mineral which undergoes both self-oxidation and air oxidation [237, 238]; while the prebiotic atmosphere did not contain oxygen, photo-activated and self-oxidation mechanisms may still have been important in vivianite and vivianite-type minerals.

The iron in vivianite is primarily present in its ferrous form. Vivianite is relatively soluble in formamide whereas inorganic ferric phosphate is nearly completely insoluble. We performed a series of formamide experiments using vivianite and ferric

phosphate as catalysts; we also employed $FeCl_3$ as a soluble Fe^{3+} control, Fe_2O_3 as an insoluble Fe^{3+} control, and $FeSO_4$ as a soluble Fe^{2+} control.

4.1.2 Small Molecule Intermediates

4.1.2.1 Small Molecules Detected on Comets, on Meteorites, and in the Interstellar Medium

A variety of small molecules may have been present on the early Earth, in addition to formamide and water; these could have formed endogenously or have been delivered by impact. A selection of molecules detected on comets, on meteorites, and in the interstellar medium is provided in Figure 4.3. Many of these small molecules would have possessed reactive chemical groups including nitriles, carboxylic acids, aldehydes, and amines; these small molecules may have enhanced or inhibited the production of nucleobases.

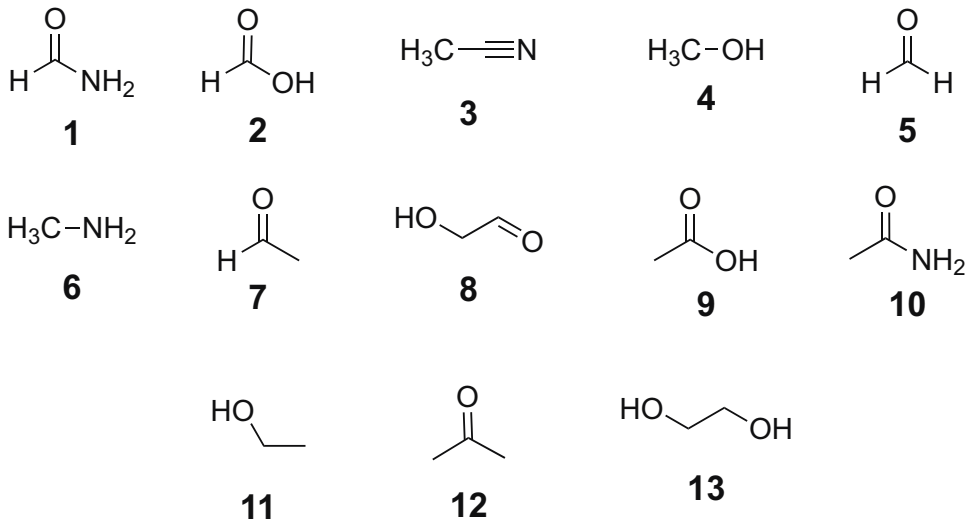

Figure 4.3 A selection of one-, two-, and three-carbon molecules detected on comets or in the interstellar medium. **1** formamide [239], **2** formic acid [240], **3** acetonitrile [241, 242], **4** methanol [241], **5** formaldehyde [243], **6** methylamine [244], **7** acetaldehyde [245, 246], **8** glycolaldehyde [247], **9** acetic acid [248], **10** acetamide [239], **11** ethanol [249], **12** acetone [250], **13** ethylene glycol [251].

We were interested in the effects these small molecules have on the production of nucleobases from formamide solutions, either via altering the acidity or basicity of the reaction medium or by participating directly in these reactions.

4.1.2.2 Role of Small Molecules in Formamide Chemistry

The roles of one-, two-, and three-carbon intermediates in formamide chemistry have not been extensively investigated, although formaldehyde has been implicated in the production of carboxylic acids from formamide [79]. However, small molecules including urea, glycolic acid, oxalic acid, and carbodiimide (Figure 4.4) have been reported as products of these reactions [36, 79].

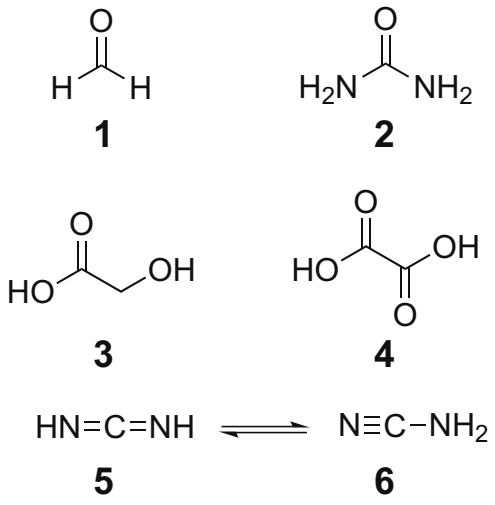

Figure 4.4 Small molecule intermediates in formamide chemistry. **1** formaldehyde has been added to formamide solutions and demonstrated to produce sugars [79]; **2** urea, **3** glycolic acid, and **4** oxalic acid have been detected as formamide reaction products [79]. **5** carbodiimide and **6** cyanamide are isomers that have also been produced in formamide reactions [79].

4.2 EXPERIMENTAL PROCEDURES

4.2.1 Materials

Solvents were HPLC grade or LC-MS grade, as appropriate, and were purchased from EMD chemicals. Inorganic salts, HPLC buffers, small molecule intermediates, and formamide were purchased from Sigma Aldrich, Fisher Scientific, or VWR, and were used as received. Bathophenanthroline and most nucleobase standards were purchased from Sigma Aldrich; 8-aminoadenine was purchased from Tocris Bioscience and 2,6-diaminopurine from Carbosynth. Water was purified using a Barnstead nanoPure system and all HPLC buffers were filtered with 0.2 μ solvent filters (Millipore) prior to use. Vivianite sourced from deposits near Moscow was purchased from Artist's Supply Source. Industrial gases were purchased from AirGas South.

4.2.2 Reaction Conditions

4.2.2.1 Phosphate Mineral Reactions

Reactions were carried out as described in Chapter 3.

4.2.2.2 Reactions with Small Molecule Intermediates

Reactions were carried out under an air atmosphere with no mineral catalyst present; 10 mM of a one-, two-, or three-carbon small molecule was dissolved in 4 mL of formamide. The chemistries of 25 different small molecules were explored; these are depicted in Figure 4.5. Reactions were then conducted as described in Chapter 3.

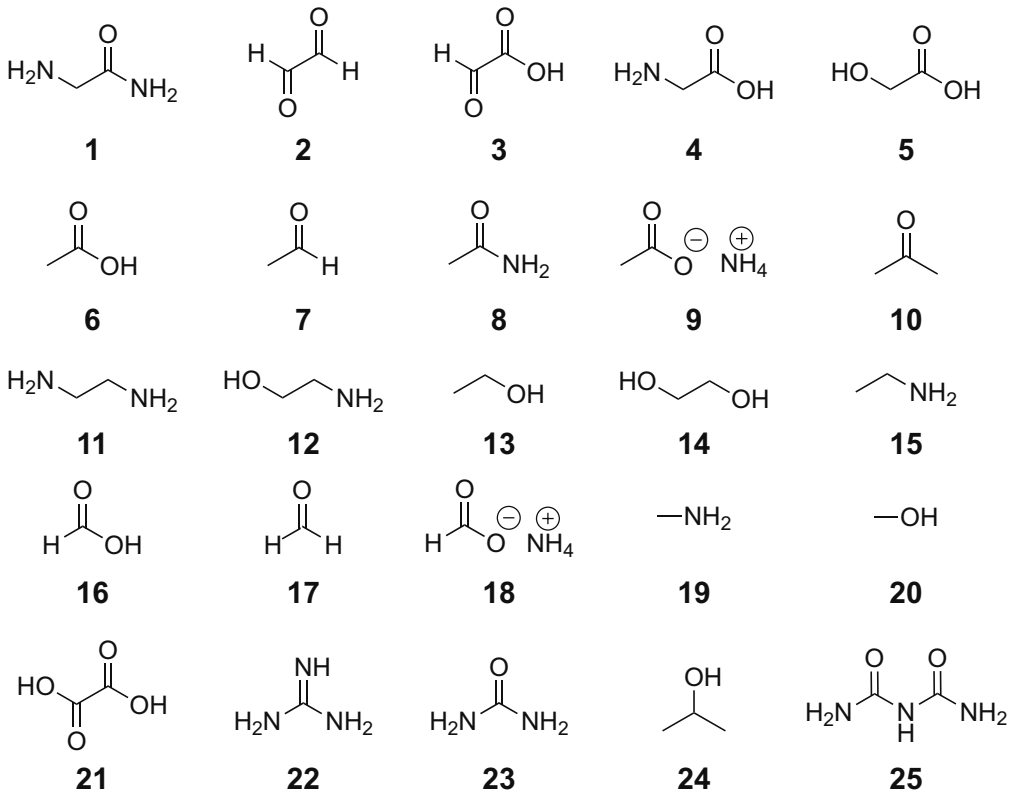

Figure 4.5 One-, two-, and three-carbon molecules spiked into formamide reactions. **1** glycinamide, **2** glyoxal, **3** glyoxylic acid, **4** glycine, **5** glycolic acid, **6** acetic acid, **7** acetaldehyde, **8** acetamide, **9** ammonium acetate, **10** acetone, **11** ethylenediamine, **12** ethanolamine, **13** ethanol, **14** ethylene glycol, **15** ethylamine, **16** formic acid, **17** formaldehyde, **18** ammonium formate, **19** methylamine, **20** methanol, **21** oxalic acid, **22** guanidine, **23** urea, **24** isopropanol, **25** biuret.

4.2.3 Sample Preparation

Aliquots of formamide reaction mixtures were dried under vacuum as described in Chapter 3.

4.2.3.1 LC-MS/MS

LC-MS/MS samples were prepared as described in Chapter 3.

4.2.3.2 HPLC

Phosphate mineral reaction samples were prepared as described in Chapter 3. Aliquots from reactions with small molecule intermediates were prepared as described in Chapter 2.

4.2.3.3 Spectrophotometric Bathophenanthroline Assay

The bathophenanthroline (Figure 4.6) assay was conducted as described in Chapter 3 on the samples which were catalyzed by iron-containing salts.

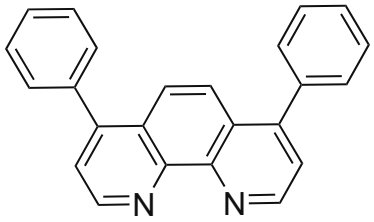

Figure 4.6 Bathophenanthroline (4,7-diphenyl-1,10-phenanthroline).

4.2.4 Instrumental Methods

4.2.4.1 LC-MS/MS

LC-MS/MS MRM analysis was carried out as described in Chapter 3; see sections 3.2.4.1.1 and 3.2.4.1.2 for details and optimal parameters for detection of nucleobase standards.

4.2.4.2 HPLC

Phosphate mineral reactions were analyzed as described in Chapter 3 using an Agilent 1100 HPLC system with binary pump and DAD. Reactions with small molecule intermediates were analyzed as described in Chapter 2 using an Agilent 1200 RRLC system with binary pump and DAD.

4.2.4.3 UV-Vis Spectroscopy

UV-Vis spectra for the bathophenanthroline assay were collected in a precision quartz cell with a 1 cm path length (Hellma) and recorded on a JASCO V515 spectrophotometer as described in Chapter 3.

4.3 RESULTS

4.3.1 Phosphate Mineral Reactions

Nine nucleobase products were detected by LC-MS/MS in reactions catalyzed by vivianite, $FePO_4$, or other soluble and insoluble iron-containing controls: adenine, 8-aminoadenine, 5-azacytosine, 2,6-diaminopurine, 2,4-diaminopyrimidine, guanine, hypoxanthine, melamine, and purine. (See Figure 3.6 for structures.)

4.3.1.1 Air Atmosphere

Purine was usually the major product in these reactions, regardless of the catalyst. In thermal-only reactions, few or no other products are observed for vivianite-catalyzed reactions, $FePO_4$-catalyzed reactions, or their respective supernatants, though traces of adenine, 5-azacytosine, and 2,4-diaminopyrimidine are found in the thermal-only vivianite sample. $FeCl_3$ and Fe_2O_3 thermal-only reaction samples generate measurable amounts of the triazines 5-azacytosine and melamine, in addition to purine, but few other products (Figures 4.7 and 4.8).

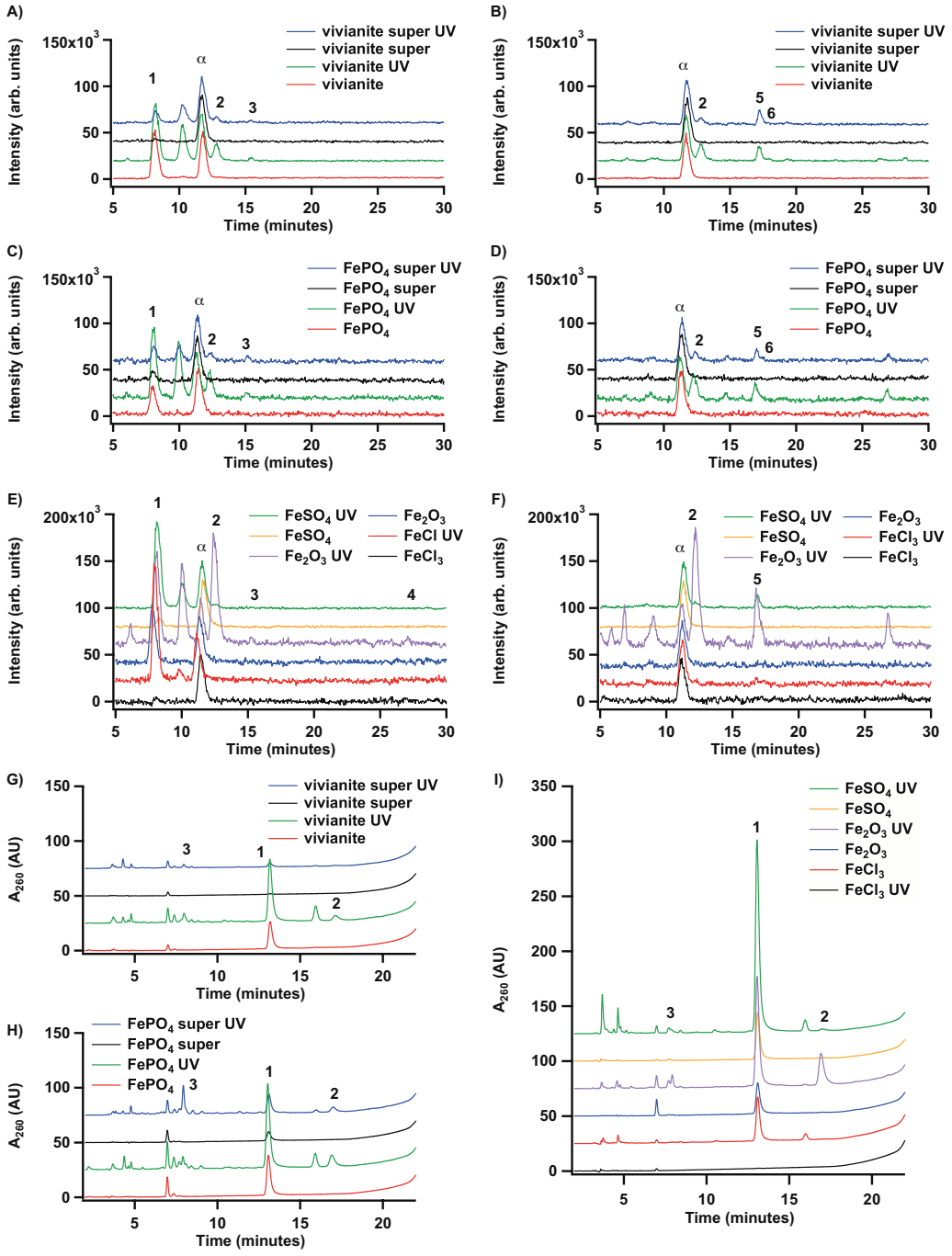

Figure 4.7 (A-F) LC-MS/MS and (G-I) HPLC chromatograms of formamide reactions conducted under an air atmosphere. Selected peaks are labeled: **1** purine, **2** adenine, **3** hypoxanthine, **4** guanine**, 5** 5-azacytosine, **6** melamine, **α** 2-aminopurine (internal standard). (A) and (B) represent two injections of the same sample but monitoring of different MS/MS transitions; this is also true of (C) and (D) and of (E) and (F). Note that (A-F) represent combined selected ion chromatograms and minor products may not be visible here, but can be seen in individual ion channels. See text and color-coded keys on individual panels for experimental details.

Reactions conducted under an air atmosphere which were both heated and UV irradiated, however, usually display larger product diversities, with all phosphate-containing samples and their supernatants producing 5-azacytosine, adenine, melamine, and hypoxanthine; $FePO_4$-catalyzed samples also produce relatively large amounts of guanine. Of the soluble and insoluble iron-containing control catalysts, however, only Fe_2O_3 with UV irradiation produces appreciable yields and product diversity (Figures 4.7 and 4.8). With the exception of the lack of guanine produced in some samples, the results for irradiated samples mirror those of the uncatalyzed formamide reaction (Figure 3.11).

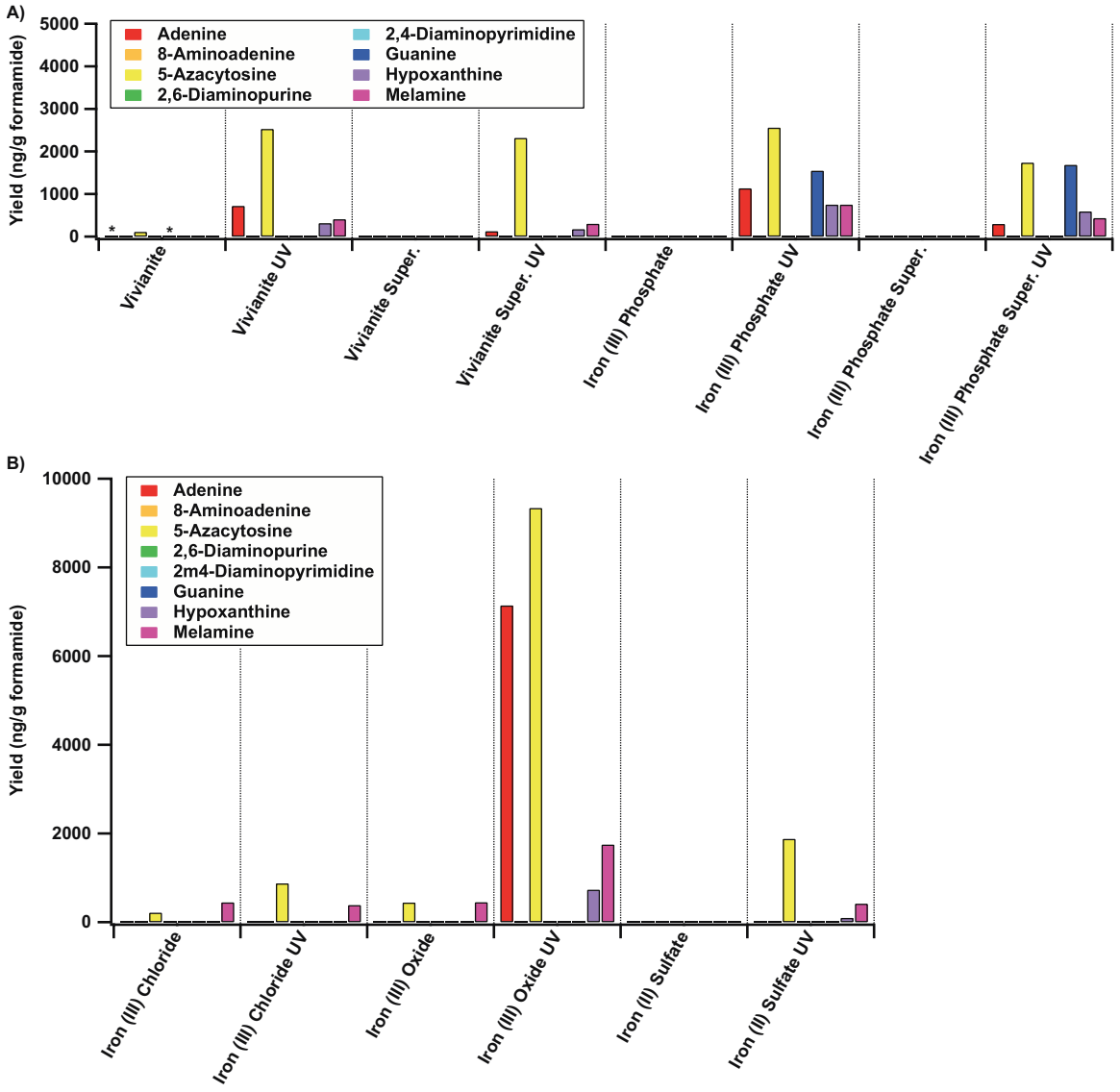

Figure 4.8 Yields of nucleobases and related compounds under an air atmosphere. In each case, 4 mL of formamide was heated at 130°C for 48 hours in the presence of 200 mg catalyst, when catalyst was present. For supernatant samples ("super." in figure legend), catalyst was mixed with formamide at room temperature, the solution was centrifuged, and the supernatant drawn off and reacted. This represents a case where a concentration of ions less than saturation are in solution and no solids are present. Some samples were irradiated with 254 nm UV light. Asterisks (*) indicate tentative identifications.

4.3.1.2 Argon Atmosphere

Under an argon atmosphere, the results prove to be quite different from our work with pyrite. Purine yields are generally highest under argon, and some guanine is still detected in the UV-irradiated vivianite supernatant sample (Figures 4.9 and 4.10). Guanine is one of the more oxidized products we monitored; the water molecules associated with vivianite in its mineral matrix may have been key to providing the oxygen required for the synthesis of guanine.

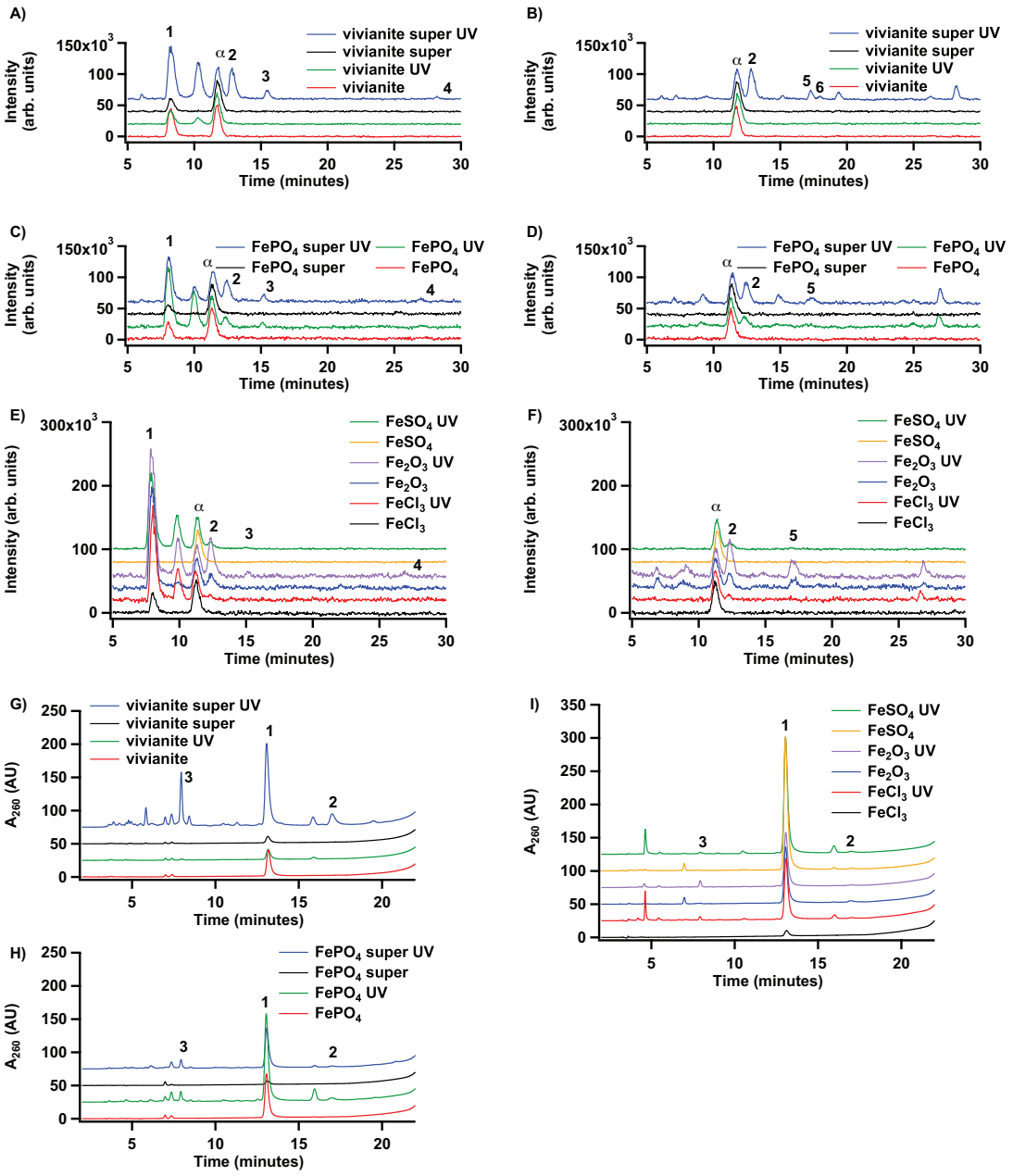

Figure 4.9 (A-F) LC-MS/MS and (G-I) HPLC chromatograms of formamide reactions conducted under an argon atmosphere. Selected peaks are labeled: **1** purine, **2** adenine, **3** hypoxanthine, **4** guanine**, 5** 5-azacytosine, **6** melamine, **α** 2-aminopurine (internal standard). (A) and (B) represent two injections of the same sample but monitoring of different MS/MS transitions; this is also true of (C) and (D) and of (E) and (F). Note that (A-F) represent combined selected ion chromatograms and minor products may not be visible here, but can be seen in individual ion channels. See text and color-coded keys on individual panels for experimental details.

As was seen with the pyrite samples in Chapter 3, under an argon atmosphere, 5-azacytosine yields are reduced in favor of the more biologically-relevant adenine and hypoxanthine, at least for vivianite and iron (III) phosphate samples (Figures 4.9 and 4.10). Small amounts of 2,4-diaminopyrimidine – the only pyrimidine detected in these experiments – are observed under several conditions, and appreciable amounts of 8-aminoadenine are seen in $FePO_4$ supernatant and Fe_2O_3 samples which have been subjected to UV exposure; 2,6-diaminopurine is also observed in the UV-exposed Fe_2O_3 sample (Figures 4.9 and 4.10). $FeCl_3$ and $FeSO_4$ remain catalytically-inactive, even under photochemical conditions (Figures 4.9 and 4.10).

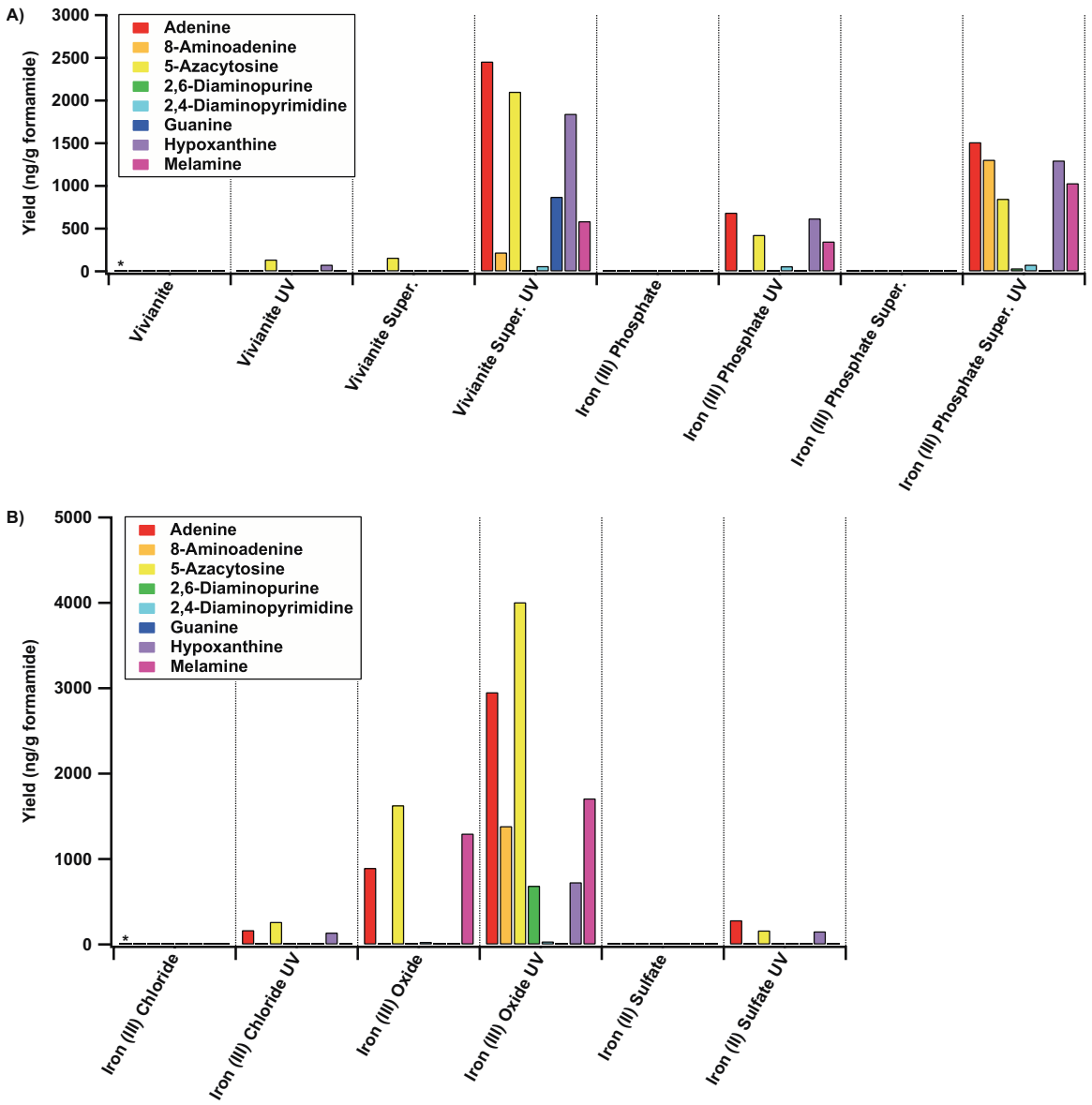

Figure 4.10 Yields of nucleobases and related compounds under an argon atmosphere. In each case, 4 mL of formamide was heated at 130°C for 48 hours in the presence of 200 mg catalyst, when catalyst was present. For supernatant samples ("super." in figure legend), catalyst was mixed with formamide at room temperature, the solution was centrifuged, and the supernatant drawn off and reacted. This represents a case where a concentration of ions less than saturation are in solution and no solids are present. Some samples were irradiated with 254 nm UV light. Asterisks (*) indicate tentative identifications.

4.3.1.3 Carbon Dioxide Atmosphere

As seen with the pyrite work presented in Chapter 3, a carbon dioxide atmosphere was the most conducive to production of nitrogen heterocycles from formamide solutions. Thermal-only reactions with vivianite produce measurable amounts of 5-azacytosine and 2,4-diaminopyrimidine, though $FePO_4$ thermal reactions produce few or no products (Figures 4.11 and 4.12). $FeCl_3$ and $FeSO_4$ are also generally inactive as catalysts under a CO_2 atmosphere (Figures 4.11 and 4.12).

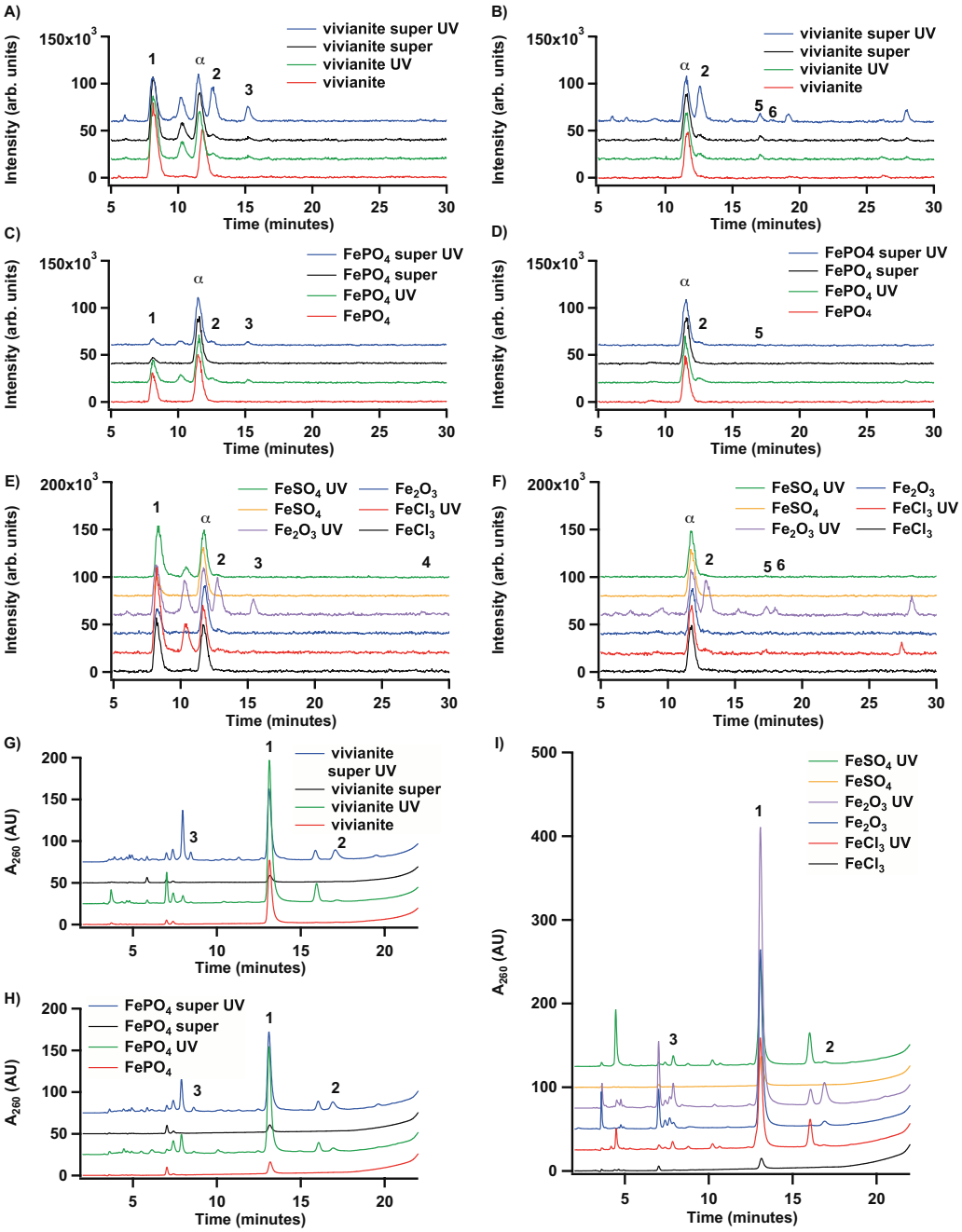

Figure 4.11 (A-F) LC-MS/MS and (G-I) HPLC chromatograms of formamide reactions conducted under a carbon dioxide atmosphere. Selected peaks are labeled: **1** purine, **2** adenine, **3** hypoxanthine, **4** guanine**, 5** 5-azacytosine, **6** melamine, **α** 2-aminopurine (internal standard). (A) and (B) represent two injections of the same sample but monitoring of different MS/MS transitions; this is also true of (C) and (D) and of (E) and (F). Note that (A-F) represent combined selected ion chromatograms and minor products may not be visible here, but can be seen in individual ion channels. See text and color-coded keys on individual panels for experimental details.

The most productive reaction conditions are UV-irradiated, with both the vivianite supernatant and Fe_2O_3 samples producing large amounts of adenine, guanine, and hypoxanthine, as well as 5-azacytosine, melamine, and 8-aminoadenine (Figures 4.11 and 4.12). While the UV-exposed vivianite sample also generates adenine and hypoxanthine, yields are much smaller than in the supernatant sample (Figures 4.11 and 4.12).

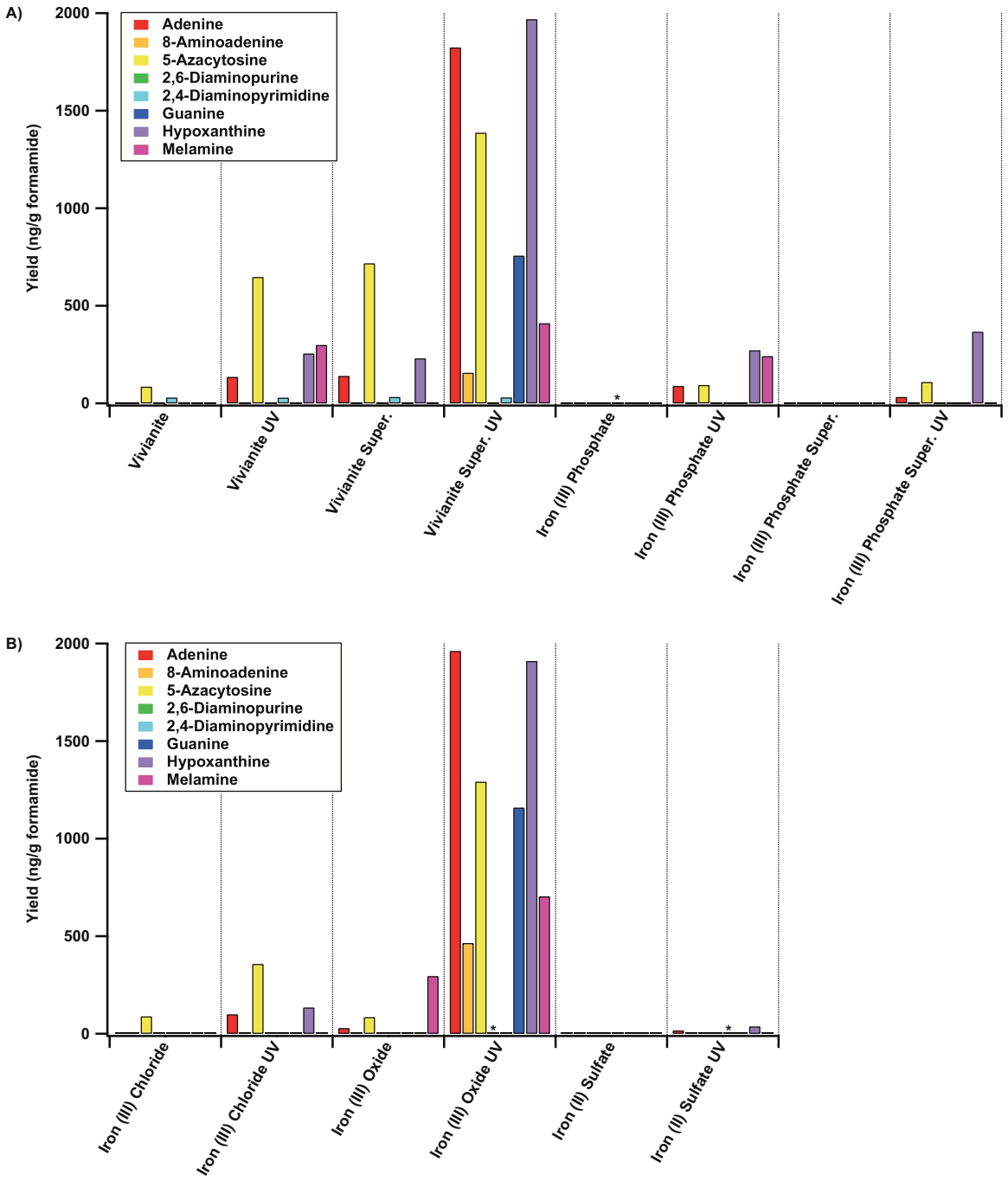

Figure 4.12 Yields of nucleobases and related compounds under a carbon dioxide atmosphere. In each case, 4 mL of formamide was heated at 130°C for 48 hours in the presence of 200 mg catalyst, when catalyst was present. For supernatant samples ("super." in figure legend), catalyst was mixed with formamide at room temperature, the solution was centrifuged, and the supernatant drawn off and reacted. This represents a case where a concentration of ions less than saturation are in solution and no solids are present. Some samples were irradiated with 254 nm UV light. Asterisks (*) indicate tentative identifications.

4.3.1.4 Cation Controls

Reactions were also performed with phosphate salts containing metals other than iron, and with the $CaCO_3$ and $Na_4P_2O_7$ salts first explored in the work in Chapter 2. Products previously reported for calcium carbonate and sodium pyrophosphate samples (guanine, adenine, hypoxanthine, and purine) were again observed; these reactions also produced small amounts of 8-aminoadenine, melamine, and 5-azacytosine when UV irradiated (Figures 4.13 and 4.14). $Ca_3(PO_4)_2$ and Na_3PO_4 samples generally produced the same products as $CaCO_3$ and $Na_4P_2O_7$ under photochemical conditions, though with slightly different yields (Figures 4.13 and 4.14).

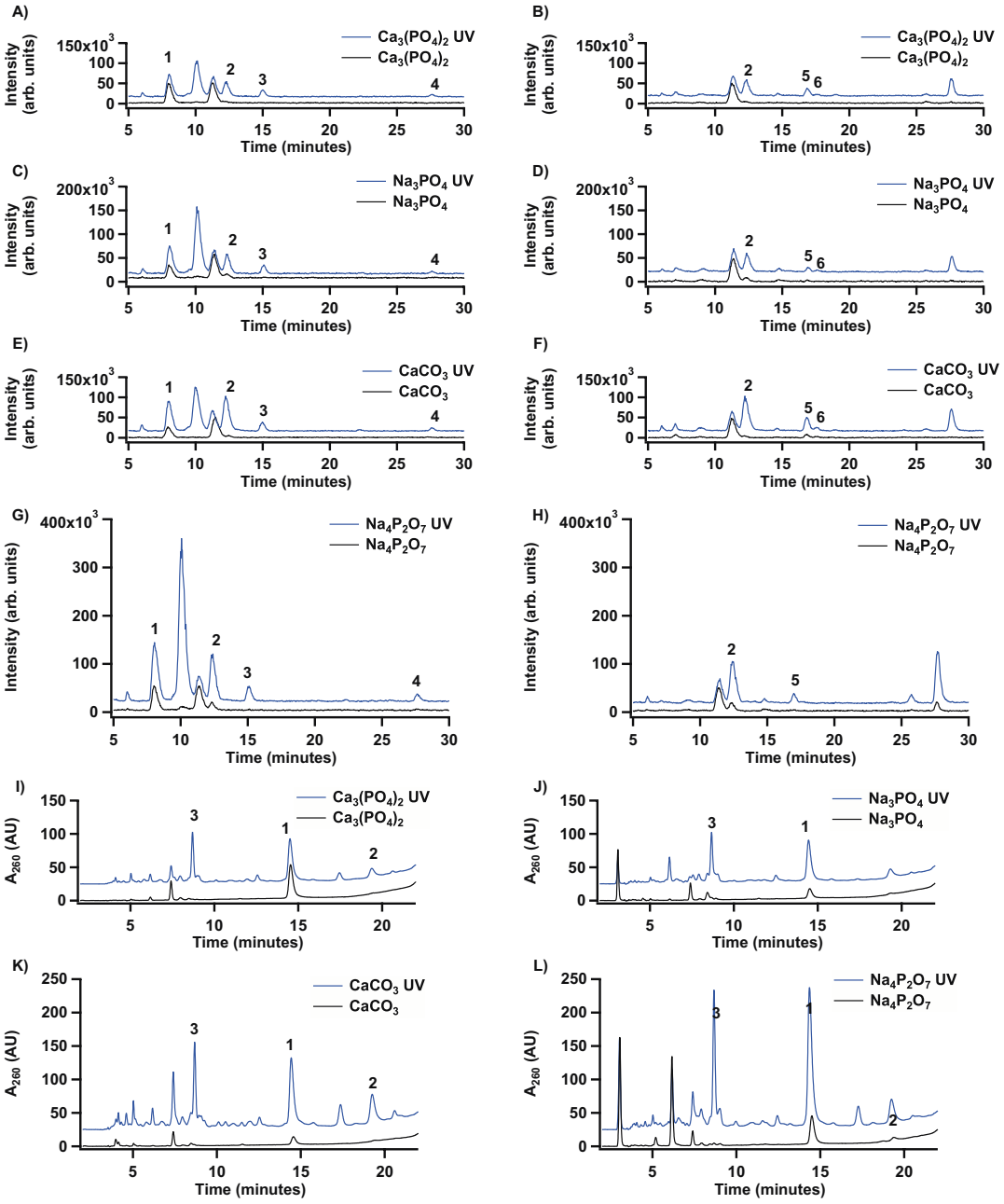

Figure 4.13 (A-H) LC-MS/MS and (I-L) HPLC chromatograms of formamide reactions conducted under an air atmosphere. Selected peaks are labeled: **1** purine, **2** adenine, **3** hypoxanthine, **4** guanine**, 5** 5-azacytosine, **6** melamine, **α** 2-aminopurine (internal standard). (A) and (B) represent two injections of the same sample but monitoring of different MS/MS transitions; this is also true of (C) and (D), of (E) and (F), and of (G) and (H). Note that (A-H) represent combined selected ion chromatograms and minor products may not be visible here, but can be seen in individual ion channels. See text and color-coded keys on individual panels for experimental details.

Additionally, HPLC analyses of the UV-irradiated samples indicate the presence of a number of unidentified compounds (Figure 4.13).

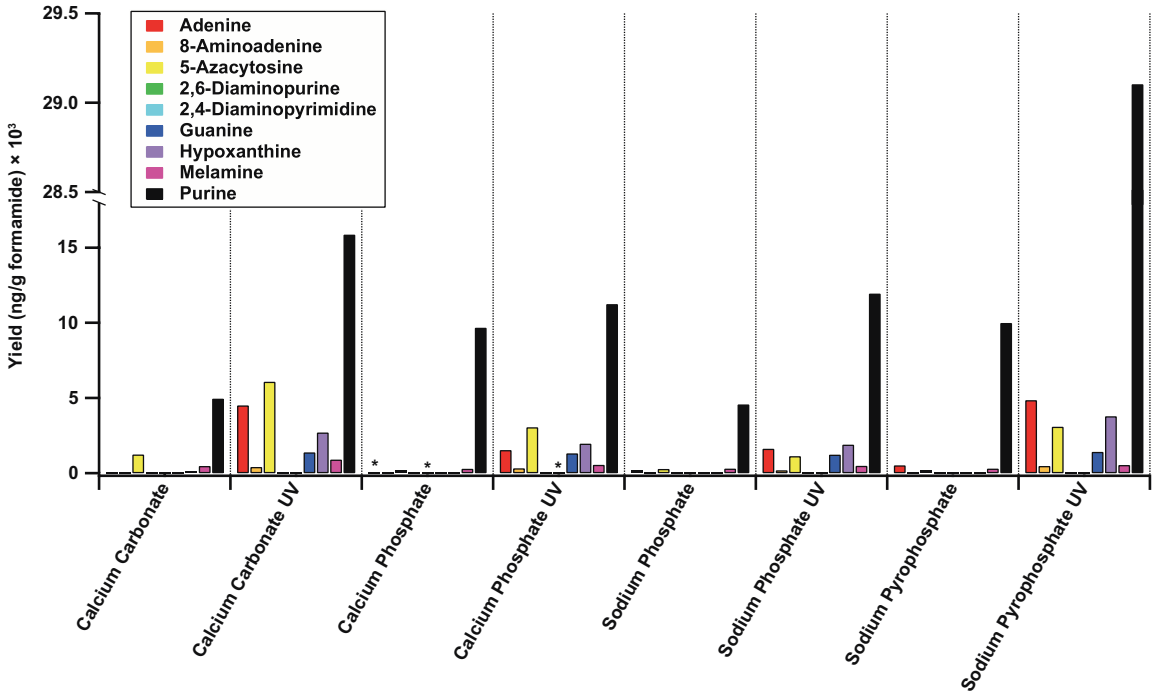

Figure 4.14 Yields of nine nucleobase products and related compounds in the presence of sodium and calcium salts. In each case, 4 mL of formamide was heated at 130°C for 48 hours in the presence of 200 mg catalyst, when catalyst was present. Some samples were irradiated with 254 nm UV light. Asterisks (*) indicate tentative identifications.

4.3.1.5 Purine Yields

Interestingly, unsubstituted purine yields were highest under an argon atmosphere, and somewhat reduced under air and carbon dioxide atmospheres, with iron oxide samples producing the highest yields of unsubstituted purine under both thermal only and combined thermal and photochemical conditions; $FePO_4$ supernatant samples produced the smallest amounts of purine under oxidizing atmospheres (Figure 4.15).

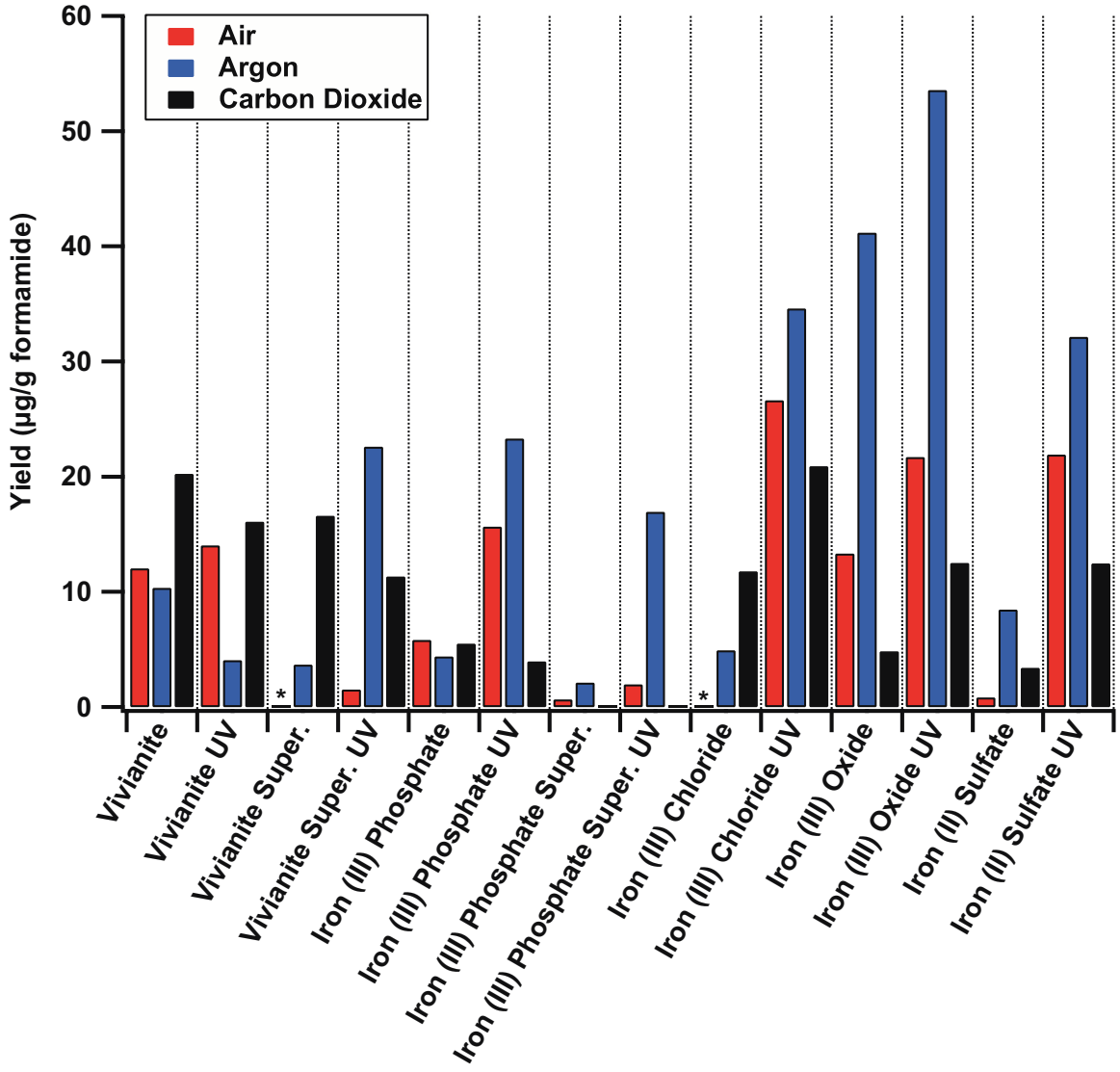

Figure 4.15 Purine yields under various reaction conditions. In each case, 4 mL of formamide was heated at 130°C for 48 hours in the presence of 200 mg catalyst, when catalyst was present. For supernatant samples ("super." in figure legend), catalyst was mixed with formamide at room temperature, the solution was centrifuged, and the supernatant drawn off and reacted. This represents a case where a concentration of ions less than saturation are in solution and no solids are present. Some samples were irradiated with 254 nm UV light. Asterisks (*) indicate tentative identifications.

4.3.1.6 Solubility of Iron Compounds

We also performed the bathophenanthroline assay on vivianite and the iron salt controls to determine the solubilities of these compounds in formamide over time and to assess the effects of Fe^{2+} and Fe^{3+} ions on the outcomes of formamide reactions. No iron (II) could be detected in the $FePO_4$ sample, and only trace amounts of iron (II) and iron (III) could be detected in the Fe_2O_3 sample, even after 48 hours at 130°C (Figure 4.16). Appreciable amounts of both cations are observed in $FeCl_3$ and $FeSO_4$ samples, and even vivianite is somewhat soluble in formamide, with a steady increase in both iron (II) and iron (III) over time (Figure 4.16).

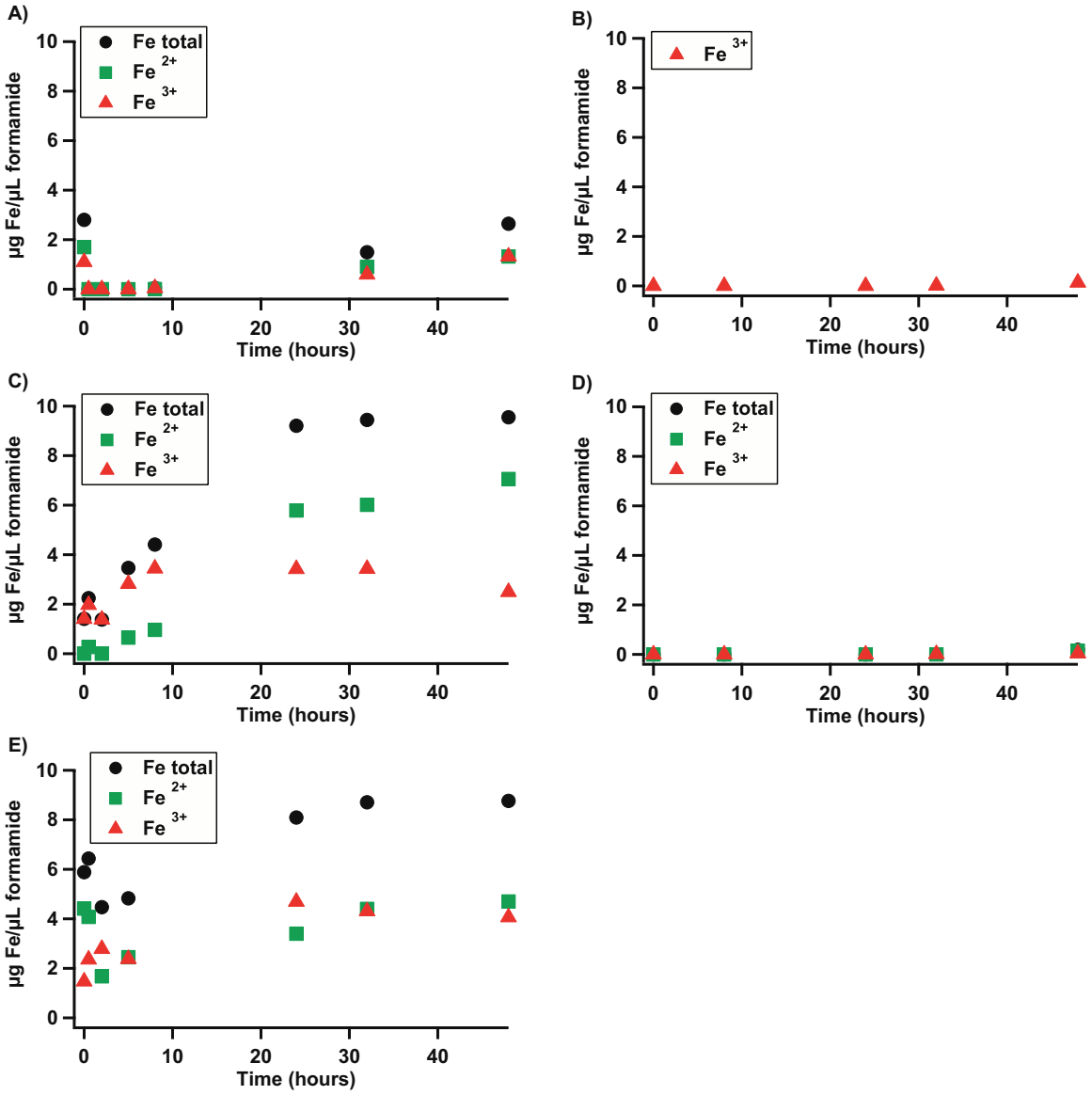

Figure 4.16 Ferric and ferrous iron in formamide reaction solutions, as determined by the bathophenanthroline spectrophotometric assay. (A) Vivianite, (B) ferric phosphate, (C) ferric chloride, (D) ferric oxide, and (E) ferrous sulfate. No ferrous iron was detected in the formamide/ferric phosphate solution at any point during the experiment.

4.3.2 Reactions with Small Molecule Intermediates

As expected, thermal-only reactions spiked with 10 mM small molecule intermediates are less productive than UV-irradiated reactions (Figure 4.17). Among non-irradiated reactions, glycinamide appears to generate the largest diversity of products; UV-irradiated reactions generally produce similar suites of products, though acetamide- and guanidine-spiked reactions, qualitatively, produce the highest number of chromatographic peaks, and ethylenediamine the lowest (Figure 4.17).

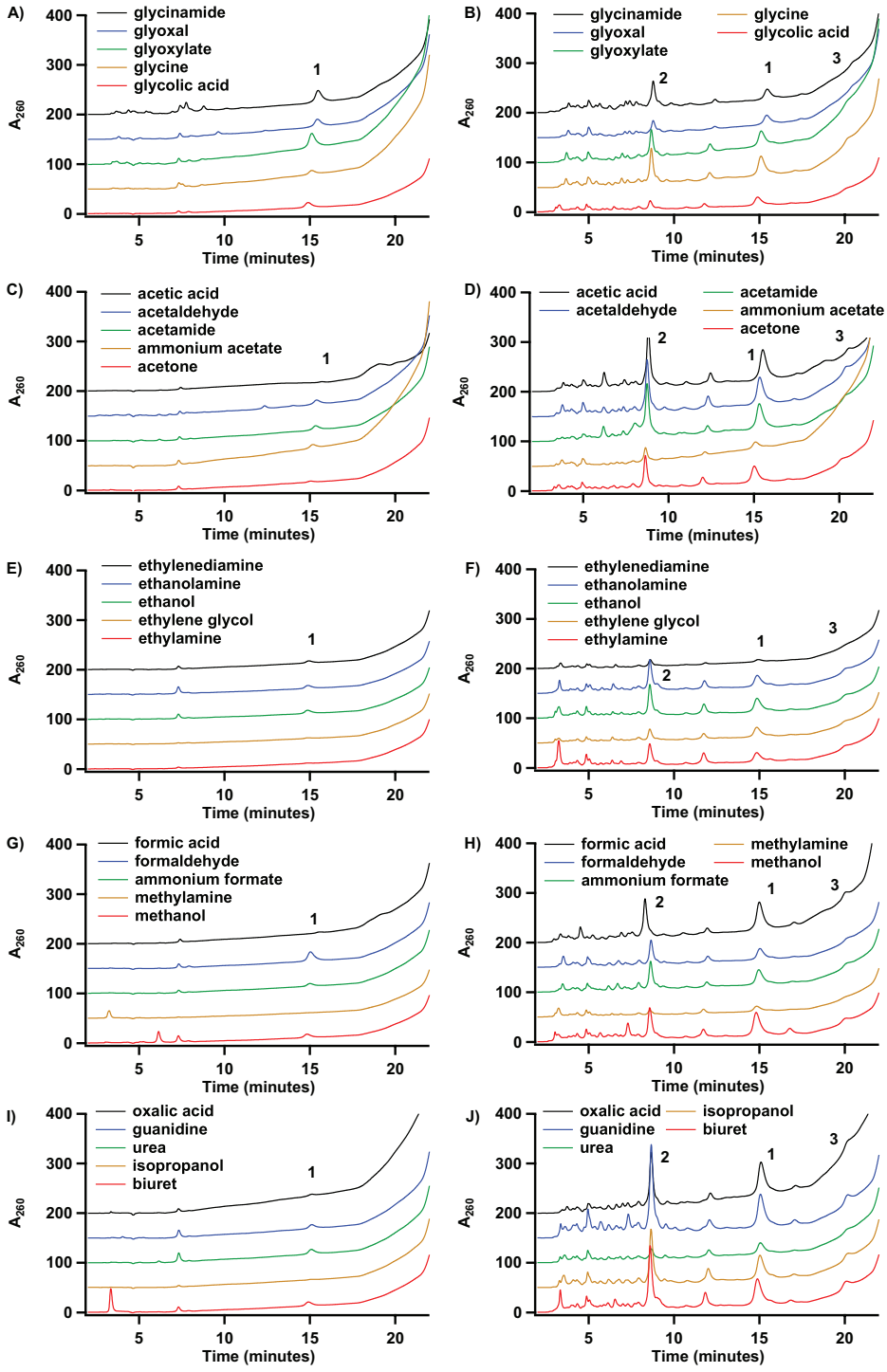

Figure 4.17 HPLC chromatograms of products of formamide reactions spiked with various small molecule intermediates. (A), (C), (E), (G), and (I) represent the products of thermal-only reactions, while (B), (D), (F), (H), and (J) represent the products of combined thermal and photochemical reactions. **1** purine, **2** hypoxanthine, **3** adenine.

4.4 DISCUSSION

4.4.1 Solubility of Iron and the Production of Nucleobases

It appears that neither soluble iron (II) nor soluble iron (III), as represented by $FeSO_4$ and $FeCl_3$, respectively, is particularly catalytically-active in formamide reactions, though it may also be the case that soluble iron is catalyzing the photodegradation of small organics such as the nucleobases and nucleobase analogs discussed here. Solution-phase chemistry may still be important, as many of the sodium- and calcium-containing phosphate salts have at least limited solubility in formamide and produce varied suites of products, including several of biological relevance. Insoluble iron (III) can be quite productive, as observed with UV-irradiated iron (III) oxide solutions, particularly for those reactions performed under a carbon dioxide atmosphere, but it can also be non-productive, as evidenced by our results with iron (III) phosphate.

4.4.2 The Roles of Cations and Anions

In these reactions, surface chemistry appears to be quite important, especially with the iron oxide samples. However, solution phase chemistry is also clearly active. Group I and II metals, represented in these experiments by sodium and calcium, respectively, may be less likely to serve as photocatalysts for the synthesis of nucleobases than transition metals such as iron, but also may be less destructive of product compounds. From the sodium and calcium series of reactions, it appears that sodium is more catalytically-active than calcium, and that anion reactivity, as measured by yield and diversity of products, follows the trend pyrophosphate > phosphate > carbonate; however, further experiments would be needed to apply these results more generally.

4.4.3 Adhesion of Products to Surfaces Versus Destruction of Products

One result that appears anomalous is the much lower yield and diversity of products generated when solid vivianite mineral is present in UV-irradiated reactions, as opposed to only dissolved vivianite in a supernatant sample. Two possible explanations suggest themselves: first, that small organics such as nucleobases may adhere to the vivianite mineral surface and be overlooked in LC-MS/MS analysis of the liquid phases of reaction mixtures, or second, that vivianite is acting as a catalyst in a photo-Fenton-type reaction [252-255] and destroying most of the purines, pyrimidines, and triazines as they are produced. The adhesion of products to a mineral surface could be beneficial as it could help to concentrate them, thereby enhancing further reactions such as nucleoside formation. The destruction of products, on the other hand, would be problematic; if significant amounts of destruction occurred, it would suggest that life likely did not originate in the presence of vivianite or similar minerals.

4.4.4 Small Molecule Intermediates

In most cases, it appears that the addition of small molecule intermediates does not significantly alter the product profile of non-mineral-catalyzed formamide reactions. In itself, this information is useful, as it tells us that the presence of various one-, two-, and three-carbon species is unlikely to interfere with the production of nucleobases, that formamide chemistry works even in a rather complex model prebiotic milieu. With some small molecules, such as acetamide and guanidine, reactivity may be slightly enhanced, especially under conditions of UV irradiation.

4.5 CONCLUDING REMARKS

Several additional experiments could be performed to help us further understand formamide chemistry in the presence of phosphate minerals. First, the use of alternative phosphate minerals as catalysts in formamide reactions may be instructive, particularly since vivianite may have been quite rare on the primitive Earth. Second, performing $CaCO_3$, $Na_4P_2O_7$, Na_3PO_4, and $Ca_3(PO_4)_2$ reactions under argon and carbon dioxide atmospheres will expand our understanding of the possible roles of cations and anions in prebiotic solution-phase formamide chemistry. A systematic investigation of different anion/cation pairs, both in their inorganic salt forms and in mineral matrices, might also be instructive. Third, to determine whether products are being lost from the UV-exposed reaction catalyzed by solid vivianite, multiple lines of evidence will be needed. Possibilities for elucidating an answer to the question of adsorption versus destruction include direct mass spec detection of products from the solid mineral surface, theoretical calculations of the favorability of the interaction between vivianite and various reaction products, and liquid chromatographic analysis of aliquots removed from reactions over time, both with and without initial spikes of products, to determine the change in levels of products after heating and irradiation. Additionally, performing phosphate mineral reactions with additional wavelengths of light may help distinguish mineral photochemistry from formamide photochemistry.

Finally, it could be worthwhile to perform some of the reactions with spiked small molecule intermediates in the presence of mineral catalysts, as many of the small molecules of interest in these experiments have one or more functional groups which could chelate metal ions in solution, providing access to additional chemistries. Using

isotopically-labeled small molecules would inform us as to whether the spiked intermediates are truly intermediates and are being incorporated into the final products or whether they play a different role. Certainly, it could also be useful and informative to characterize and identify additional unknowns present in any or all formamide reaction mixtures, not just those discussed in this chapter.

CHAPTER 5

MOLECULAR RECOGNITION OF WATSON-CRICK-LIKE PURINE-PURINE BASE PAIRS*

5.1 INTRODUCTION

Proto-RNAs likely shared some characteristics with DNA and RNA, including semi-conservative replication and base pairing mediated by hydrogen bonds, as well as alternating sugars and linkage units in the backbone. Many plausibly prebiotic syntheses of nucleobases and nucleobase-like compounds have been reported; however, in addition to the canonical adenine, cytosine, uracil, and guanine found in RNA today, numerous other purine and pyrimidine bases have been identified as products of these reactions. Thus, the earliest nucleic acids may very well have contained bases which are not found in DNA or RNA – or indeed, anywhere in life – today.

Some of these noncanonical bases have been preserved; for example, inosine is frequently found in the wobble position of the anticodon loop of transfer RNAs, as well as in some messenger RNAs [256]. Purine-purine base pairs have also been demonstrated to exist in the ribosome and in tRNAs [257, 258]. These nonstandard pairings could be vestiges of a proto-RNA world.

5.1.1 An All-Purine Nucleic Acid Precursor

The inability of pyrimidine nucleotides to polymerize in an enzyme-free system on a polypurine template [259] led Wächtershäuser in 1988 to propose an all-purine nucleic acid precursor. It has long been known that pyrimidine bases stack poorly in

* The results presented in this chapter were previously published in Buckley, R. *et al*, *ChemBioChem*, **2011**, *12*, 2155-2158.

solution, and this phenomenon may contribute to the suppressed polymerization [45, 145]. In Wächtershäuser's model, adenine and guanine bases are glycosylated at the normal N(9) position and pair with N(3) glycosylated xanthine and isoguanine, respectively (Figure 5.1).

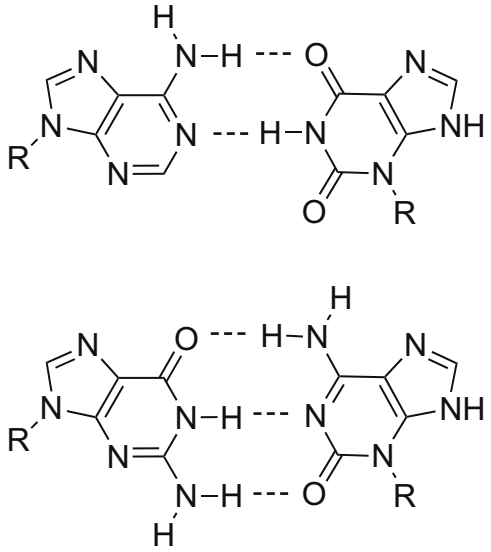

Figure 5.1 Purine-purine base pairs proposed by Wächtershäuser [45]. Top: adenine·xanthine pair with xanthine glycosylated at the N(3) position and adenine glycosylated at the usual N(9) position. Bottom: guanine·isoguanine pair with isoguanine glycosylated at the N(3) position and guanine glycosylated at the usual N(9) position. R represents a sugar unit.

While this model has yet to be tested experimentally, Wächtershäuser notes that existing biosynthetic pathways could lead to N(3) glycosylated purines and suggests that there is a primitive sequence of metabolic reactions that could branch to generate both N(3) and N(9) glycosylated species. Wächtershäuser further supports his notion of the antiquity of purine-based evolution by noting that many coenzymes are derivatives of purines, but not of pyrimidines [45].

5.1.2 Antiparallel Purine-Purine Duplexes

Recently, two research groups have demonstrated the formation of antiparallel homopurine duplexes in which every nucleobase is attached to its sugar – in these cases, DNA – at the 9 position [43, 44] (Figure 5.2).

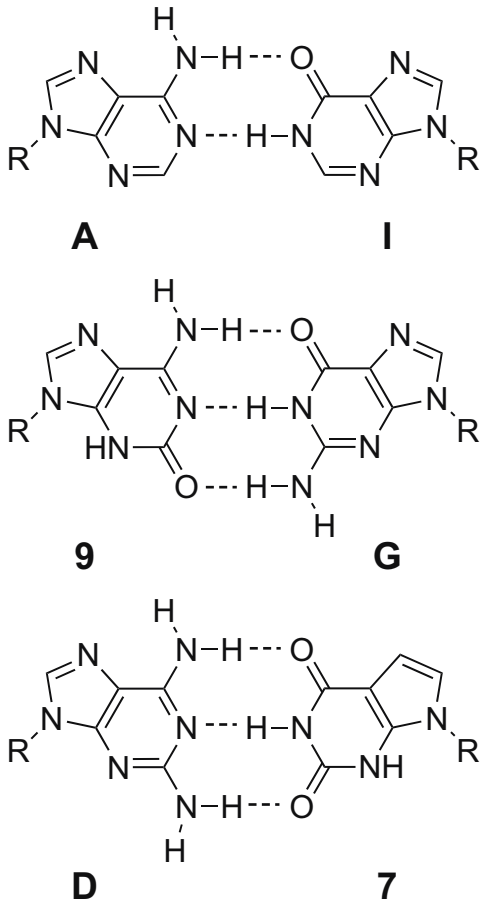

Figure 5.2 Purine-purine base pairs investigated by Battersby and coworkers (top and middle) [44] and by Heuberger and Switzer (middle and bottom) [43]. **A** adenosine, **I** inosine, **9** isoguanosine, **G** guanosine, **D** diaminopurine, **7** 7-deazaxanthosine. R represents a sugar unit.

Battersby and colleagues reported sequence-specific assembly of antiparallel all-purine DNA duplexes consisting of the adenosine·inosine and guanosine·isoguanosine

base pairs, in a 1:1 ratio. However, their homopurine duplexes exhibited significantly lower melting temperatures than equivalent Watson-Crick duplexes containing mixed purine and pyrimidine bases [44]. Heuberger and Switzer, meanwhile, incorporated a 2,6-diaminopurine nucleotide and the purine analogue 7-deazaxanthosine into their sequences; this base pair (Figure 5.2) includes an additional hydrogen bond and duplexes incorporating it exhibit thermal stabilities comparable to their purine-pyrimidine counterparts [43]. We have chosen a DNA backbone for our study, though it should be noted that sequences which vary only in their backbones may have significantly different thermodynamics of DNA/drug interaction [260, 261].

5.1.3 Molecular Recognition

Molecular recognition can be thought of as a specific, noncovalent interaction between two molecules; it may be affected by multiple energetic contributions including, but not limited to, aromatic stacking, hydrogen bonding, and cation-π interactions [262], or a combination of these. Molecular recognition is critical in the field of drug design, where scientists gather information on the structures of biological targets and use this knowledge as a guide to develop pharmaceutical agents with maximum efficacy and minimum side effects. Indeed, many small-molecule drugs which are currently in clinical use against a variety of diseases recognize and bind to specific DNA and/or RNA sequences (see Section 1.7).

5.1.4 Small Molecule Interactions with All-Purine Nucleic Acid Structures

Intercalators (Section 1.7.1.1) stack between nucleobases [263-265] and lead to increases in the melting temperature of nucleic acid duplexes [260]. Since nucleic acid duplexes containing adenosine·inosine and guanosine·isoguanosine base pairs are

plausibly prebiotic and less thermally stable than their Watson-Crick counterparts [44], and since it has been hypothesized that a molecular midwife resembling an intercalator may have assisted in the assembly of the first nucleic acids [188], we searched for a midwife analogue from among known drugs that would selectively stabilize homopurine duplexes. We hypothesized that such a molecule would have a large, crescent-shaped aromatic stacking surface, complementary in size and shape to a purine-purine base pair. Studies of molecular recognition of polypurine structures are not without precedent; small molecules that bind to polyadenylic acid and to G-quadruplexes have been reported.

5.1.4.1 Polyadenylic Acid

Several classes of ligands (Figure 5.3) target polyadenylic acid, including nucleobase and nucleoside analogs such as formycin B and 7-methylxanthine [266-268] and natural products such as berberine and palmatine [157, 269, 270]. Further, some synthetic azacyanine dyes [271] and other molecules have been reported to induce a self-structure in poly(A) at neutral pH [157]. Many, though not all, of these molecules possess large, crescent-shaped stacking surfaces which may interact favorably with the various A·A self-structures [156, 158].

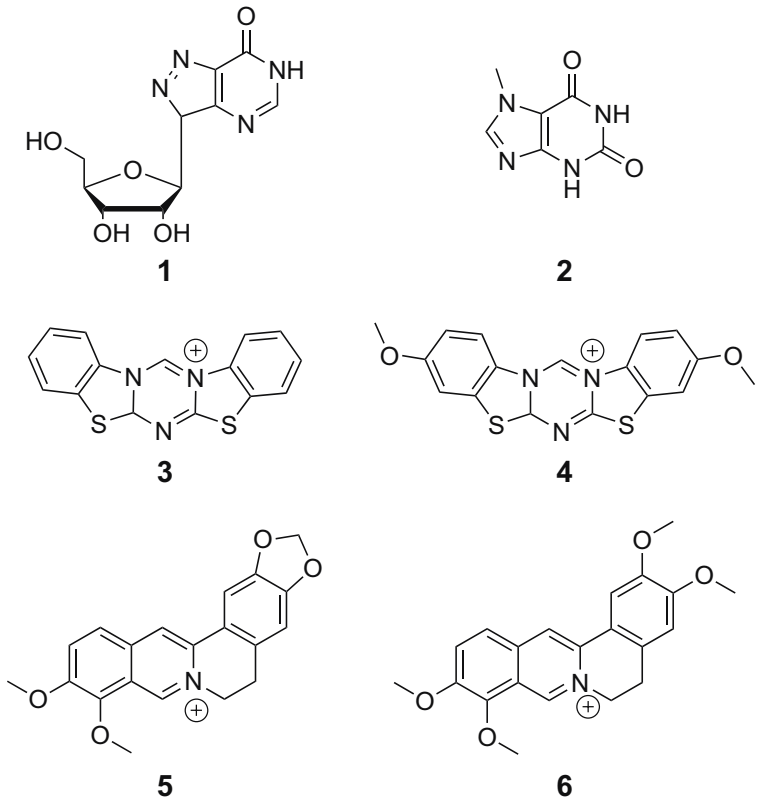

Figure 5.3 Selected poly(A) ligands. **1** formycin B (allopurinol riboside), **2** 7-methylxanthine, **3** azacyanine4, **4** azacyanine5, **5** berberine, **6** palmatine.

5.1.4.2 G-Quadruplexes

G-quadruplexes present larger surface areas than A·A mismatches; consequently, we might expect quadruplex ligands to have even larger surface areas than polyadenylic acid ligands. In fact, porphyrin analogs such as TMPyP4 and NMM, and the mixed thiazoline/oxazole molecule telomestatin, exhibit favorable binding properties to various G-quadruplexes; telomestatin has been sown to induce quadruplex formation, even in the absence of a suitable cation [272].

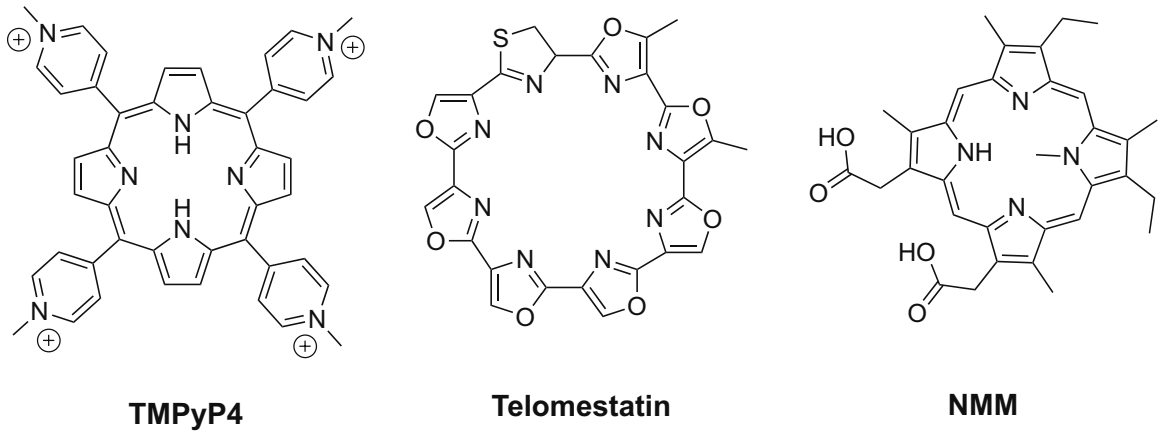

Figure 5.4 Selected G-quadruplex ligands. NMM is believed to be anionic at physiological pH [272].

5.1.5 Small Molecules Used in This Study

Small molecules with a variety of aromatic architectures were selected for this study of molecular recognition of purine-purine base pairs. These fell into three groups: (1) molecules with smaller surface areas which are known to intercalate natural DNA, (2) larger, crescent-shaped molecules which are expected to stack well with purine-purine base pairs, and (3) molecules which intercalate natural DNA but which also have prosthetic groups that interact with the minor groove. With the latter group, it was hypothesized that tight binding to homopurine DNA duplexes would not be reported due to expected differences in minor groove width.

5.1.5.1 Proflavine

Proflavine (Figure 5.5) is an acridine dye and known mutagen with two exocyclic amino groups. Its nucleic acid-binding properties have been thoroughly investigated. In 1961, Lerman reported that proflavine and other acridines enhanced viscosity and reduced the sedimentation coefficients of DNA solutions, consistent with an intercalative mode of binding [263]; later, a crystal structure of proflavine with an RNA dinucleotide

[273] and NMR structure of proflavine with a 2′,5′-RNA duplex [261] demonstrated that proflavine's main binding mode with these nucleic acids is intercalation. Some additional minor binding modes have also been reported [274].

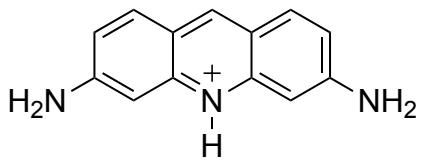

Figure 5.5 Chemical structure of proflavine.

We hypothesized that proflavine would bind to both homopurine and Watson-Crick DNA. Its stacking surface, with only three rings, may be closer to the size of a purine-pyrimidine base pair, and thus proflavine would be expected to interact more favorably with a standard DNA duplex than with an all-purine duplex.

5.1.5.2 Ethidium

Ethidium (Figure 5.6), also referred to as homidium, has seen long use in veterinary medicine to treat trypanosomosis in cattle [275]. This molecule was formerly a popular stain for nucleic acids in gel electrophoresis, though it has largely been replaced by non-mutagenic stains; it exhibits fluorescence enhancement upon intercalation of DNA and RNA [187, 276-278].

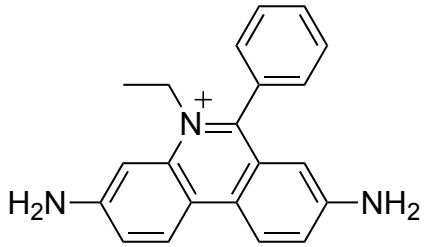

Figure 5.6 Chemical structure of ethidium.

Ethidium was also hypothesized to bind to both homopurine and Watson-Crick DNA duplexes. Ethidium, like proflavine, has three rings in its stacking surface; its pendant ring projects into the minor groove [187]. We also expected ethidium to interact more favorably with Watson-Crick DNA than with homopurine DNA.

5.1.5.3 Ellipticine

Ellipticine (Figure 5.7) is a plant alkaloid with a slightly larger surface area than proflavine; it binds much more tightly to DNA than does proflavine, also by intercalation [279-281]. This drug and several of its derivatives are potent anticancer agents [281, 282].

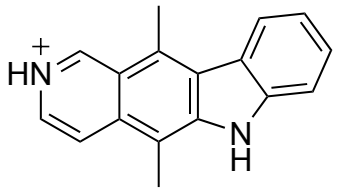

Figure 5.7 Chemical structure of ellipticine.

Because ellipticine is one of the tightest known binders to natural DNA [281], we hypothesized that it would bind selectively to a mixed purine-pyrimidine duplex;

however, the slightly larger surface area and crescent shape of ellipticine suggested that this molecule might also exhibit favorable interactions with homopurine DNA.

5.1.5.4 Coralyne

Coralyne (Figure 5.8) is a member of the family of isoquinoline alkaloids that also includes berberine, palmatine, and sanguinarine [283]. Its binding properties to poly(A), including its ability to induce a poly(A) self-structure have been thoroughly characterized [157, 158, 269, 270, 284]; this molecule has also been demonstrated to disproportionate a poly(dT)·poly(dA) duplex into a poly(dT)·poly(dA)·poly(dT) triplex and a poly(dA) single strand [285].

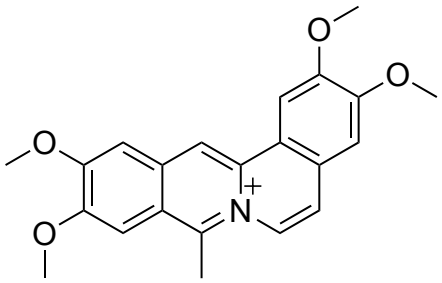

Figure 5.8 Chemical structure of coralyne.

Coralyne was selected for this study because of its large, crescent-shaped stacking surface and its demonstrated ability to interact with larger base-paired structures. While this molecule demonstrates some binding to mixed purine·pyrimidine strands [270, 283], we hypothesized that it would interact more favorably with an all-purine DNA duplex.

5.1.5.5 Azacyanine3

Azacyanine3 (aza3, Figure 5.9) is a member of a class of small molecules originally designed as chloride-selective ion channel modulators for potential use in

patients with cystic fibrosis [271]. Like coralyne, this molecule possesses an extended aromatic stacking surface and a crescent shape and is able to induce a poly(A) self-structure at neutral pH [157].

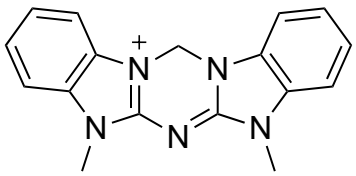

Figure 5.9 Chemical structure of azacyanine3 (aza3).

Aza3 binds with micromolar or better affinity to DNA with the human telomeric G-quadruplex sequence (d(TTGGG(TTAGGG)$_3$A) by stacking on the terminal G-tetrad of this sequence [286]. Because of this previously-demonstrated interaction with a large aromatic surface, it was expected that aza3 would bind preferentially to our homopurine DNA duplex.

5.1.5.6 Daunomycin

Daunomycin, also known as daunorubicin, is an anthracycline natural product which is used to treat various forms of leukemia [186]. This drug is known to intercalate DNA with its long axis perpendicular to the long axis of the base pairs; its daunosamine moiety projects into the minor groove, where it makes specific contacts with residues [186], preferring GC-rich sequences as binding sites [270, 287].

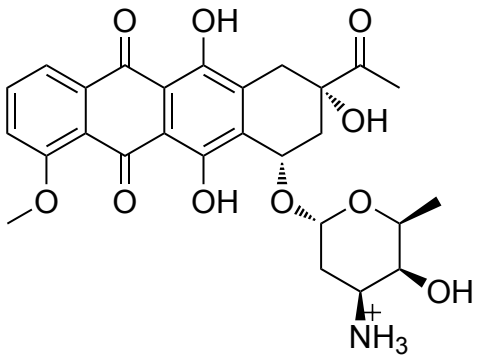

Figure 5.10 Chemical structure of daunomycin (daunorubicin).

The minor groove of a homopurine DNA double helix is expected to be wider than that of a standard Watson-Crick helix; it is thus unlikely that daunomycin could make the same stabilizing contacts in the minor groove with an all-purine sequence, as opposed to a mixed purine-pyrimidine sequence. We therefore hypothesized that daunomycin would bind poorly to our purine-purine duplex.

5.1.5.7 Actinomycin D

Actinomycin D, like daunomycin, is a natural product isolated from *Streptomyces* bacteria; this drug is used in the treatment of some cancers because of its cytotoxic properties [185, 288]. Actinomycin D's chromophore intercalates at GpC sites, with its depsipentapeptide rings making extensive contacts in the minor groove [185]. The actinomycin D/DNA interaction requires a longer time to come to equilibrium than do other drug/DNA interactions, presumably because the perturbed cyclic peptides must reorient upon binding [288].

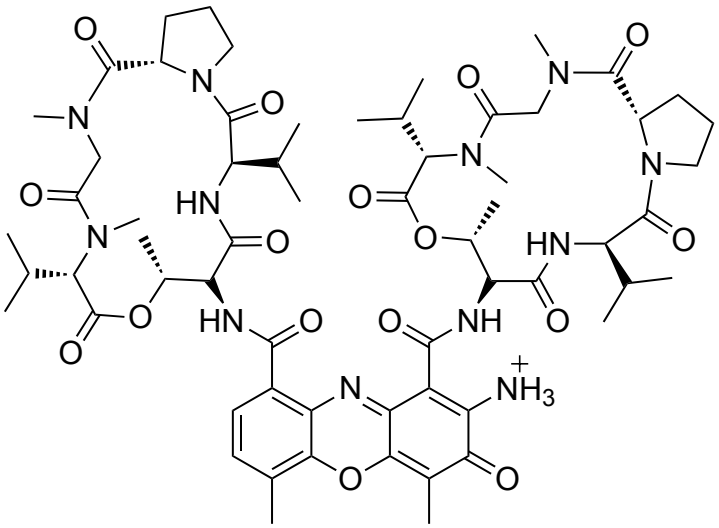

Figure 5.11 Chemical structure of actinomycin D.

Given the expected differences in the groove geometries of natural DNA and purine-purine DNA, we expected that actinomycin D, like daunomycin, would exhibit reduced binding to our all-purine sequence.

5.1.6 Spectroscopic Methods for the Investigation of Small-Molecule Nucleic-Acid Interactions

While X-ray crystallography and NMR are the ultimate arbiters of DNA/drug interaction at the atomic level, crystallization and/or interpretation of NMR spectra is not possible or practical for all nucleic acid-small molecule complexes of interest. Other spectroscopic techniques exist, however, which can provide evidence supporting a particular mode of binding, or insight into the nature of an interaction.

5.1.6.1 UV-Visible Spectrophotometry

UV-Visible (UV-Vis) spectrophotometry is a widely applicable type of absorbance spectroscopy and is particularly useful in the investigation of DNA/small

molecule interactions because the DNA bases and all of the small molecules under investigation in this study possess extensive π electron systems.

It is generally assumed that molecules in solution at room temperature are in the ground electronic and vibrational states [289]; when these molecules are excited with UV and/or visible light, a number of electronic transitions become possible, though in practice, ππ* transitions are the only ones with enough intensity to be discerned [289]. Due to collisions of excited molecules in the solution phase, these transitions are usually observed as broad bands [289]. DNA/drug interactions involve changes to ππ stacking and evidence for these interactions can be seen in the UV-Vis spectra of these complexes.

5.1.6.1.1 Thermal Denaturation

In a thermal denaturation, or melting, experiment, UV-Vis spectra are obtained at incrementally higher temperatures. When a DNA duplex unwinds or a DNA-drug complex dissociates, ππ stacking is disrupted and overall absorbance at a given wavelength increases; this is known as hyperchromicity. The melting temperature (T_m) is defined as the temperature at which half of a given duplex (or complex) is assembled and half is dissociated.

T_m values offer insight into the thermal stabilities of DNA duplexes and DNA-drug complexes. If a small molecule binds tightly to a DNA duplex, but not particularly well to a dissociated single strand, an increased T_m will be observed for the duplex in the presence of that small molecule.

5.1.6.1.2 Binding Stoichiometry

Using Job's method of continuous variation, the binding stoichiometry of a drug to DNA of a given sequence can be determined. In this method, concentrations of small

molecule and DNA duplex are continually altered, while the total number of moles of drug plus binding sites remains the same [290, 291]. The absorbance spectrum at each of these mole ratios of drug:binding sites is recorded. These spectra are normalized, generally to small molecule concentration, and the wavelength of maximum absorbance or a ratio of absorbances is plotted. An extremum on the plot indicates the ratio of drug:binding sites at which maximum complex formation occurs; curvature of the plotted lines leading to this extremum indicates incomplete complex formation [292].

5.1.6.1.3 Binding Mode

While the final determination of binding mode can only be determined via the solution of an NMR or crystal structure of a drug/DNA complex, experimental evidence consistent with one or another mode of binding can be obtained. A binding stoichiometry of one small molecule for every two base pairs is consistent with the nearest neighbor exclusion principle and thus with intercalation (see Section 1.7.1.1). Further evidence for intercalation can be observed in the UV-Vis spectra of the free drug and of the drug/DNA complex. The ππ stacking of the DNA duplex is altered by the insertion of an intercalator; this manifests as a bathochromic (red) shift in the wavelength of maximum absorbance as well as in an observed hypochromic effect in the drug's absorbance bands [292].

5.1.6.1.4 Binding Affinity

UV-Vis dilutions can also be used to determine the binding affinity of a small molecule for a particular type of DNA or other nucleic acid duplex [157]. This affinity can be expressed as an equilibrium or association constant (K_a) for the following process:

$$[\text{drug}] + [\text{binding sites}] \rightleftharpoons [\text{complex}]$$

$$K_a = [\text{complex}]/([\text{drug}]\times[\text{binding sites}])$$

In a dilution experiment, the drug:binding sites ratio remains constant, though the total concentrations of drug and of DNA duplex (and therefore of binding sites), decrease. As buffer solution is added to the sample, the complex begins to dissociate; UV-Vis spectra are obtained at each concentration and then previously described properties such as bathochromic shift (see section 5.1.6.1.3) can be used to determine the fraction of drug bound according to the method of Qu and Chaires [293].

5.1.6.2 Circular Dichroism Spectroscopy

Chiral molecules absorb right- and left-circularly-polarized light differently; the difference between these absorbances is known as circular dichroism (CD). CD spectroscopy is commonly employed in the study of chromophore-containing biological macromolecules such as DNA and proteins, due to the presence of homochiral sugars and amino acids in these molecules, respectively.

Many drug molecules are achiral and have no optical activity but will exhibit an induced band upon complexation with DNA; this band is located in the region where the drug absorbs UV and/or visible light and may be positive, negative, or biphasic [294]. Induced bands in CD are a direct result of the chiral environment of DNA that surrounds the small molecule and may be used as an indication that drug/DNA interaction is occurring. Many of the same properties and quantities that can be measured by UV-Vis spectroscopy (binding stoichiometry, binding affinity, T_m) can also be measured by CD.

5.1.6.3 Fluorescence Spectroscopy

Fluorescence occurs when an electronically excited molecule emits radiation and passes back to the ground state or other lower excited state [295]. Fluorescence emission is accompanied by a Stokes shift, which is emission at a longer wavelength than at which absorption occurs; this is due to the presence of molecules in excited vibrational states [295].

Natural DNA does not exhibit strong fluorescence, but many small molecules exhibit changes in fluorescence upon binding to DNA. Ethidium, for example, shows an enhancement in fluorescence emission when complexed with DNA [278]; excited ethidium can have a long fluorescence lifetime but can be quenched when it collides with water and undergoes non-radiative decay processes [277]. Protection from solvent, e.g. by association with a nucleic acid duplex, thus leads to enhanced fluorescence [277]. By contrast, other intercalators such as proflavine exhibit decreases in fluorescence upon binding to nucleic acids [260].

5.2 EXPERIMENTAL PROCEDURES

5.2.1 Materials

Proflavine hemisulfate, coralyne chloride, actinomycin D, and daunomycin were purchased from Sigma-Aldrich and used as received. Ethidium bromide was purchased from Fisher Scientific and ellipticine from EMD Biosciences; both were used as received. Azacyanine3 was synthesized according to a previously published procedure [271]. All other chemicals and reagents were purchased from Sigma-Aldrich, VWR, or Fisher Scientific.

5.2.2 Sample Preparation

5.2.2.1 Oligonucleotide Synthesis and Purification

Purine-pyrimidine DNA with the self-complementary sequence 5′-CAGTAGCTACTG-3′ (hereafter referred to as **WC**) was purchased from Integrated DNA Technologies and used as received. Phosphoramidites and reagents for DNA synthesis were purchased from Glen Research and/or ChemGenes. Purine-purine DNA with the self-complementary sequence 5′-9AGIAG9IA9IG-3′ (hereafter referred to as **Pu·Pu**), where 9 represents isoguanine residues and I represents inosine residues, was synthesized using an automated Expedite DNA synthesizer via standard protocols but with 2.5% dichloroacetic acid used for deblocking steps. The dimethoxytrityl protecting groups were left intact on the 5′ residues.

Pu·Pu was deprotected using concentrated aqueous ammonia at 55°C overnight. The ammonia was evaporated under vacuum and the remaining solutions were frozen and lyophilized in the presence of a small amount of Tris base to prevent deblocking of the 5′-dimethoxytrityl groups. These samples were resuspended in water and purified by reverse phase HPLC. Fractions containing full-length trityl-on product were pooled and solvent removed by evaporation under vacuum.

HPLC-purified oligonucleotides were resuspended in 20% acetic acid (v/v) for one hour at room temperature to deblock the 5′ residues, then adjusted to pH 7 with triethylamine and desalted using C18 Sep Pak solid phase extraction cartridges (Waters). Fractions containing DNA were pooled and lyophilized. Stoichiometric triethylammonium associated with the phosphate groups was removed by running the

sample over an AG-50W X8 cation exchange column (Bio Rad) that had been previously equilibrated with LiOH solution.

5.2.2.2 Stock Solution Preparation

The concentrations of stock solutions of small molecules were determined spectrophotometrically using the following extinction coefficients: proflavine ε_{444} = 38 900 M^{-1} cm^{-1} ; ethidium ε_{482} = 5500 M^{-1} cm^{-1}; ellipticine ε_{295} = 60 000 M^{-1} cm^{-1}; daunomycin ε_{480} = 11 500 M^{-1} cm^{-1}; actinomycin D ε_{441} = 26 400 M^{-1} cm^{-1}; azacyanine3 ε_{343} = 47 700 M^{-1} cm^{-1}; coralyne ε_{420} = 14 500 M^{-1} cm^{-1}. The stock solution of actinomycin D was prepared in methanol and the stock solution of ellipticine was prepared in DMSO and aliquots of both were dried prior to adding these molecules to experimental samples. All other stock solutions were prepared in water.

The extinction coefficient of **WC** was provided by the manufacturer (116 100 M^{-1} cm^{-1}). The extinction coefficient for **Pu·Pu** was estimated as 106 380 M^{-1} cm^{-1}, or 90% of the sum of the extinction coefficients at 260 nm (M^{-1} cm^{-1}) for the component deoxynucleosides: 15 400, A; 11 700, G; 4600, 9; 7700, I [44]. Concentrations of stock solutions of **Pu·Pu** and **WC** were also determined spectrophotometrically.

5.2.2.3 Sample Conditions

All samples were prepared in 10 mM lithium-cacodylate buffer at pH 6.0 with 50 mM LiCl (also the buffer for dilution experiments). Unless otherwise noted, samples were 100 μM in DNA strand. For thermal denaturation experiments, small molecules, when used, were present in 300 μM concentrations; for dilution, fluorescence, and binding mode studies, small molecules were present in 100 μM concentrations.

5.2.3 Instrumentation and Computational Details

5.2.3.1 High Pressure Liquid Chromatography

Purification of **Pu·Pu** was conducted on an Agilent 1100 Series HPLC using an ODS semi-preparatory column (Phenomenex, 10.0 × 250 mm, 5μ particle size). Separation of full-length oligonucleotide from truncation products was achieved with a gradient of 30 mM NH_4OAc (A) and acetonitrile (B), requiring 5-30% B (0-10 min), 30-100% B (10-30 min), hold at 100% B (30-40 min), 100-5% B (40-42 min), hold at 5% B (42-50 min), flow rate 4.6 mL/min. HPLC buffers were filtered with 0.22μ Steritop Filters (Millipore).

5.2.3.2 UV-Vis Spectroscopy

5.2.3.2.1 Thermal Denaturation and Binding Stoichiometry

UV-Vis Job plot and thermal denaturation experiments were performed using a Hewlett Packard 8453 UV-visible diode array spectrophotometer with an Agilent 8909A Peltier temperature controller. Job plot analysis was performed at 5°C in a 1 mm path-length quartz cuvette (Jasco) with the sum of [drug] and [purine nucleobase]/4 fixed at 300 μM. Absorbance at 450 nm was normalized to proflavine concentration; for ethidium, the ratio of absorbances at 476 nm and 545 nm was monitored.

UV-Vis melting profiles were acquired from 5°C to 95°C with 1°C steps. Melting temperatures for most samples were determined by performing a least-squares fit of each spectrum as a weighted sum of the 5°C and 95°C spectra in the longer-wavelength region, where only drug absorbance occurs. For drug-free samples, this procedure was followed in the spectral regions where the DNA nucleobases absorb. For samples containing coralyne, absorbance was monitored at 437 nm and at 414 nm and the

ratio of these two curves yielded a sigmoidal melting curve with a clear transition. This sigmoidal curve was normalized to a maximum value of 1.0, representing 100% duplex assembly. For **WC** DNA, absorbance was monitored at 280 nm and normalized to a maximum value of 1.0. For ellipticine with **WC** DNA, absorbance was monitored at 260 nm and normalized to a maximum value of 1.0. Melting experiments consisted of two heating and cooling cycles and all melt curves were reversible and monophasic with the exceptions of the daunomycin and **Pu·Pu** system and all melting experiments involving actinomycin D. All actinomycin D melt samples exhibited considerable aggregation and precipitation of a highly-colored product.

5.2.3.2.2 Binding Mode and Binding Affinity

Dilution experiments in which a buffer and salt solution is added incrementally to a sample containing a constant ratio of two drug molecules per duplex, were conducted to determine the association constants for each small molecule to **Pu·Pu** and **WC**. These experiments were performed using an Ocean Optics USB 2000 UV-Vis spectrophotometer with light source and detector separated by approximately 11 cm and connected by a custom cable. Dilution experiments were conducted in a cold room at 4°C. Samples containing actinomycin D were allowed to equilibrate for 15 minutes prior to collecting spectra; other samples reached equilibrium within a few seconds after dilution. Each starting sample was 100 μM in DNA oligonucleotide (50 μM duplex) and 100 μM in small molecule.

Rectangular quartz cuvettes with path lengths of 1 mm, 2 mm, 5 mm, and 1 cm and cylindrical quartz cuvettes with path lengths of 2 cm, 5 cm, and 10 cm (Jasco and/or Starna Cells) were used to collect UV-Vis spectra of each sample, with shorter path

lengths used for more concentrated samples and longer path lengths used for less concentrated samples. Spectra were collected until the signal in the drug absorbance region of each spectrum dropped to 0.1 absorbance units in the cell with the longest path length (10 cm).

Five spectra were averaged for each concentration and were then background corrected and normalized to small molecule concentration. For drugs with complex spectral shapes in the longer-wavelength regions – coralyne, azacyanine3, and daunomycin – the fraction of drug bound was determined by performing a least-squares fit of the drug-absorbing region of the UV-Vis spectrum as a weighted sum of the normalized absorption spectrum of the free drug and of the drug in complex with DNA at the highest concentration studied. For drugs with single peaks in the longer-wavelength regions – proflavine, ethidium, and actinomycin D with both types of DNA and ellipticine with **WC** – the wavelength of maximum absorption was monitored. Since ellipticine's longest-wavelength absorption peak has a low extinction coefficient, the peak centered at 310 nm was tracked in these experiments. As **Pu·Pu** in duplex form has significant absorbance at 300 nm with a tail extending to 330 nm, a dilution of the purine-purine duplex was performed and these spectra were subtracted from the corresponding UV-Vis spectra of **Pu·Pu** with ellipticine. The resultant spectra possessed a peak near 310 nm which shifted during the course of the dilution experiment; the position of this peak was monitored to determine association constant in this case. A single curve representing fraction of drug bound was generated from this data, and fitted by a least-squares method to the following equation [293] which can be solved quadratically and is

derived from an expression for association constant for ligands binding to a molecule having multiple binding sites, such as DNA:

$$K_a x^2 - x(K_a S_0 + K_a D_0 + 1) + K_a S_0 D_0 = 0$$

where K_a is association constant, x is the concentration of ligand bound, S_0 is total number of binding sites, and D_0 is total number of ligand molecules in the system. Due to experimental design, the ratio of total number of binding sites was always the same fixed multiple of the total number of drug molecules in the system, allowing simplification of this equation. For an oligonucleotide duplex with 12 bases, we assumed 11 binding sites for a model intercalator. This fit was robust and the values of K_a obtained were of the same order of magnitude regardless of the number of sites; values from 6 to 13 sites were tested.

The quadratic equation was divided by D_0 to yield the fraction of total ligand in the system that was bound, rather than the concentration of bound ligand. To compensate for the fact that we cannot assume 100% ligand binding in any case, a weighting factor was incorporated into the solved quadratic equation before curve fitting. In most cases, a plateau was observed in the binding curves at high ligand concentrations, and we were confident that all ligand molecules were bound. The weighting factor in these cases had a value of 1. For drug molecules without an obvious plateau, the weighting factor was also 1 in most cases, except with actinomycin D with **Pu**·**Pu**. In this instance, the curve representing the fraction of drug bound was corrected using the weighting factor.

Spectra presented to demonstrate bathochromic and hypochromic shifts as indicators of binding mode are simply the highest concentration dilution samples, or else

samples of free drugs with concentrations matching the highest concentration bound drug samples.

5.2.3.3 Circular Dichroism Spectroscopy

Circular dichroism spectra were collected at 5°C on a Jasco J-810 CD spectropolarimeter equipped with a Peltier temperature control device. Samples were 100 μM in DNA strand and 300 μM in small molecule; a 1 mm path length cuvette (Jasco) was used.

5.2.3.4 Fluorescence Spectroscopy

Fluorescence experiments were performed on a Jobin-Yvon Horiba FluoroLog-3 spectrofluorometer in a quartz cuvette with three polished windows and a path length of 3 mm (Hellma Cells) maintained by a at 4°C. Samples containing ethidium were 100 μM in DNA oligonucleotide (50 μM duplex), if DNA was present, and 100 μM in ethidium. Fluorescence emission spectra were corrected for the wavelength-dependent properties of the emission monochromator.

Samples were excited at 507 nm where absorbance of free ethidium, ethidium with **WC**, and ethidium with **Pu·Pu** had the same absorbance value. Emission spectra were collected from 510 nm to 780 nm. **WC** and **Pu·Pu** without ethidium were also excited and emission spectra were collected the DNA alone. **WC** and **Pu·Pu** duplexes without ethidium did not exhibit emission spectra when excited at 507 nm. Exit and entrance slits were held at 2 nm and integration time at 0.15 s for all experiments.

5.2.3.5 Molecular Dynamics Simulations

Molecular dynamics simulations were performed using the AMBER11 suite of programs with the parmbsc0 modifications to the parm99 forcefield [296]. The structure

of **Pu·Pu** was generated using the nucgen module of AMBER11; two complementary homopurine strands were combined and missing atoms were added manually. The nucleic acid was neutralized by the addition of 22 Na^+ counterions and solvated in an octahedral box with a 10 Å water layer using the TIP3P model.

Geometry optimizations of the structures of inosine and isoguanosine were performed using the QCHEM package at the B3LYP/6-31G* level, followed by derivation of atomic partial charges using the R.E.D. III program [297]. Forcefield parameters were taken from the parm99 forcefield, and nonexistent angles and dihedrals were approximated from existing parameters.

Solvent minimization was performed (holding solute molecules fixed with a 500 kcal/mol-$Å^2$ force constant) in two stages: 500 steps of steepest descent and 500 steps of conjugate gradient minimization. This was repeated three times with consecutive reduction of the force constant to 100, 50, and 10 kcal/mol-$Å^2$. An unrestrained minimization was then performed for 2500 steps. The system was heated from 0 to 278 K in a single 50 ps run. Equilibration was performed in five successive 20-ps runs at constant pressure, with restraints of 50, 40, 30, 10, and 5 kcal/mol-$Å^2$, respectively. All restraints were removed for production runs.

NPT production trajectories were performed on the complex using a Langevin thermostat with periodic boundary conditions. The SHAKE algorithm was applied to constrain all covalent hydrogen bonds. A 2 fs time step was used and long-range electrostatics were computed using the particle mesh ewald (PME) summation method with a 12 Å cut-off. Coordinates were saved every 5 ps for a total of 50 ns.

WC was generated using the nucgen module of AMBER11, neutralized with 22 Na^+ counterions, and solvated with a 10 Å layer of TIP3P water. Solvent and counterions were minimized for 1000 steps, with a restraint of 500 kcal/mol-$Å^2$ on the DNA, followed by 2500 steps of minimization without restraints. A 20-ps constant volume run was used to heat the system from 0 to 300 K; a 1 ns equilibration was performed on the system, followed by a 45-ns production run. Coordinates were saved every 5 ps. The trajectories were analyzed using the ptraj module of the AMBER package. Curves+ [298] was used to provide helical parameters for individual DNA structures and over entire trajectories.

5.3 RESULTS

5.3.1 Melting Temperatures and Circular Dichroism of Drug-Free Duplexes

As expected, **Pu·Pu** melted at a lower temperature than **WC**, 20°C versus 46°C, respectively; two heating and two cooling profiles were obtained for each duplex, and all melts were reversible (Figure 5.12 A and B). Both duplexes exhibited intrinsic circular dichroism spectra at 5°C, though the average intensity of **Pu·Pu** was somewhat lower than for **WC** under identical experimental conditions (Figure 5.12 C).

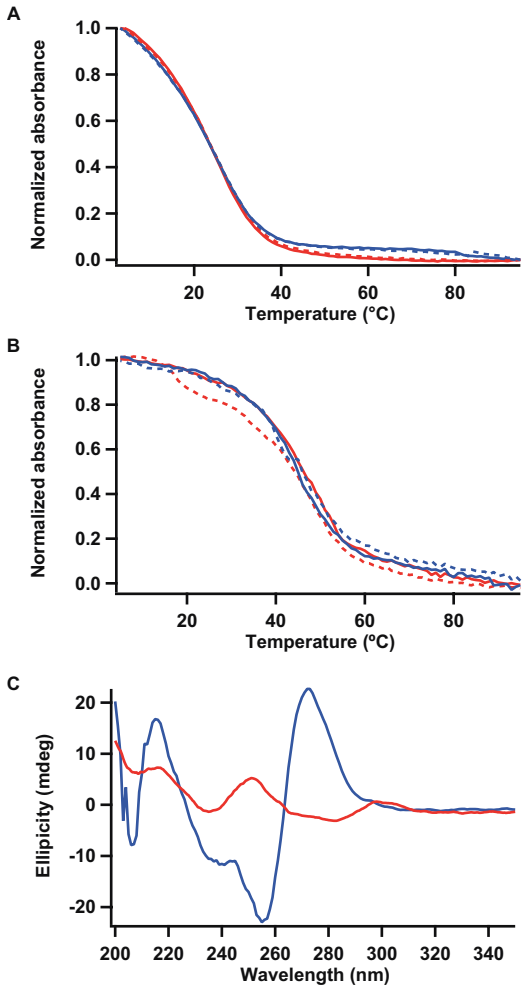

Figure 5.12 UV-Vis thermal denaturation curves for (A) **Pu·Pu** and (B) **WC**. Red lines represent heating traces, blue lines cooling traces. First heating and cooling traces are solid lines and second heating and cooling traces are dashed lines. (C) Circular dichroism spectra at 5°C for **Pu·Pu** (red) and **WC** (blue).

5.3.2 Proflavine Binding to WC and Pu·Pu

The addition of a nearest-neighbor concentration of proflavine leads to a ΔT_m of 20°C for **WC** and of 9°C for **Pu·Pu**. As with drug-free DNA, melt traces were reversible (Figure 5.13 A and B). The principal interaction of proflavine with Watson-Crick DNA is intercalation, though other binding modes such as cooperative external stacking have been reported [274]. Intercalative binding is marked by a drug:DNA ratio of one small

molecule for every two base pairs; Job plot analysis revealed this same binding stoichiometry with our **Pu·Pu** duplex (Figure 5.13 C). While an X-ray or NMR structure would need to be solved to unconditionally prove an intercalative mode of binding, our Job plot results are consistent with intercalation for proflavine and the **Pu·Pu** duplex.

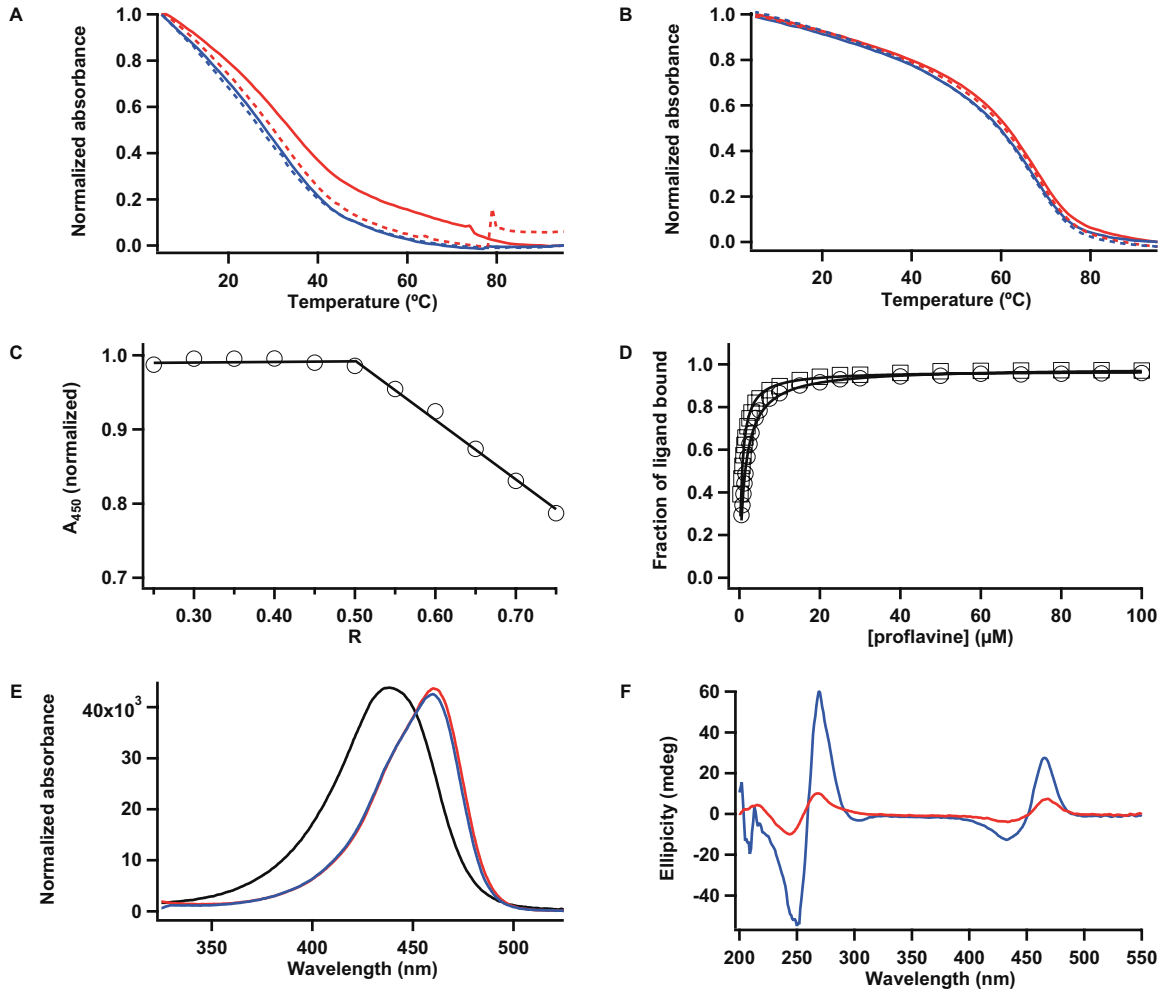

Figure 5.13 UV-Vis thermal denaturation curves for (A) **Pu·Pu** with proflavine and (B) **WC** with proflavine. Red lines represent heating traces, blue lines cooling traces. First heating and cooling traces are solid lines and second heating and cooling traces are dashed lines. (C) Job plot of proflavine with **Pu·Pu** indicating a nearest-neighbor binding stoichiometry, with R = [ligand]/([ligand]+[bp]/2). (D) Plots of the fraction of ligand bound to **WC** (open squares) and **Pu·Pu** (open circles) as a function of ligand concentration. (E) UV-Vis spectra of proflavine molecules free (black) and in the presence of **WC** (blue) or **Pu·Pu** (red). (F) Circular dichroism spectra at 5°C for **Pu·Pu** with proflavine (red) and **WC** with proflavine (blue).

A UV-Vis dilution experiment was performed to determine the respective association constants for the proflavine + **Pu·Pu** and proflavine + **WC** systems; these experiments were completed at 4°C to ensure maximum assembly of the **Pu·Pu** duplex. Performing these studies at other temperatures would result in different K_a values. As predicted (see Section 5.1.5.1), proflavine exhibits a tighter binding to **WC** (3.3×10^5 M^{-1}) than to **Pu·Pu** (3.3×10^5 M^{-1}) (Figure 5.13 D). UV-Vis spectra for both systems and for free proflavine, obtained during the dilution experiment, were compared to obtain additional evidence for a specific mode of binding. A bathochromic (red) shift and a hypochromic effect in the long-wavelength regions of drug absorbance are consistent with an intercalative mode of binding [299]; clear bathochromic shifts of 22 nm are evident in both the proflavine + **Pu·Pu** and proflavine + **WC** systems (Figure 5.13 E).

Further evidence for proflavine binding to **Pu·Pu** and **WC** is seen in the CD spectra of both of these types of DNA in solutions containing proflavine; both exhibit induced CD bands in the region of proflavine absorbance (Figure 5.13 F). The magnitudes of these bands are different, with ICD of proflavine being more intense with **WC** than with **Pu·Pu**.

5.3.3 Ethidium Binding to WC and Pu·Pu

As was the case for proflavine, the addition of ethidium at concentrations consistent with the nearest neighbor exclusion principle to solutions of **Pu·Pu** and **WC** resulted in increases in T_m, with a ΔT_m for **Pu·Pu** of 11°C and for **WC** of 25°C. Again, all melts were reversible and all melt traces were superimposable (Figure 5.14 A and B). Binding stoichiometry was also determined using Job's method of continuous variation for the ethidium + **Pu·Pu** system and was found to be one ethidium molecule for every

two base pairs, consistent with the nearest neighbor exclusion principle of intercalation (Figure 5.14 C).

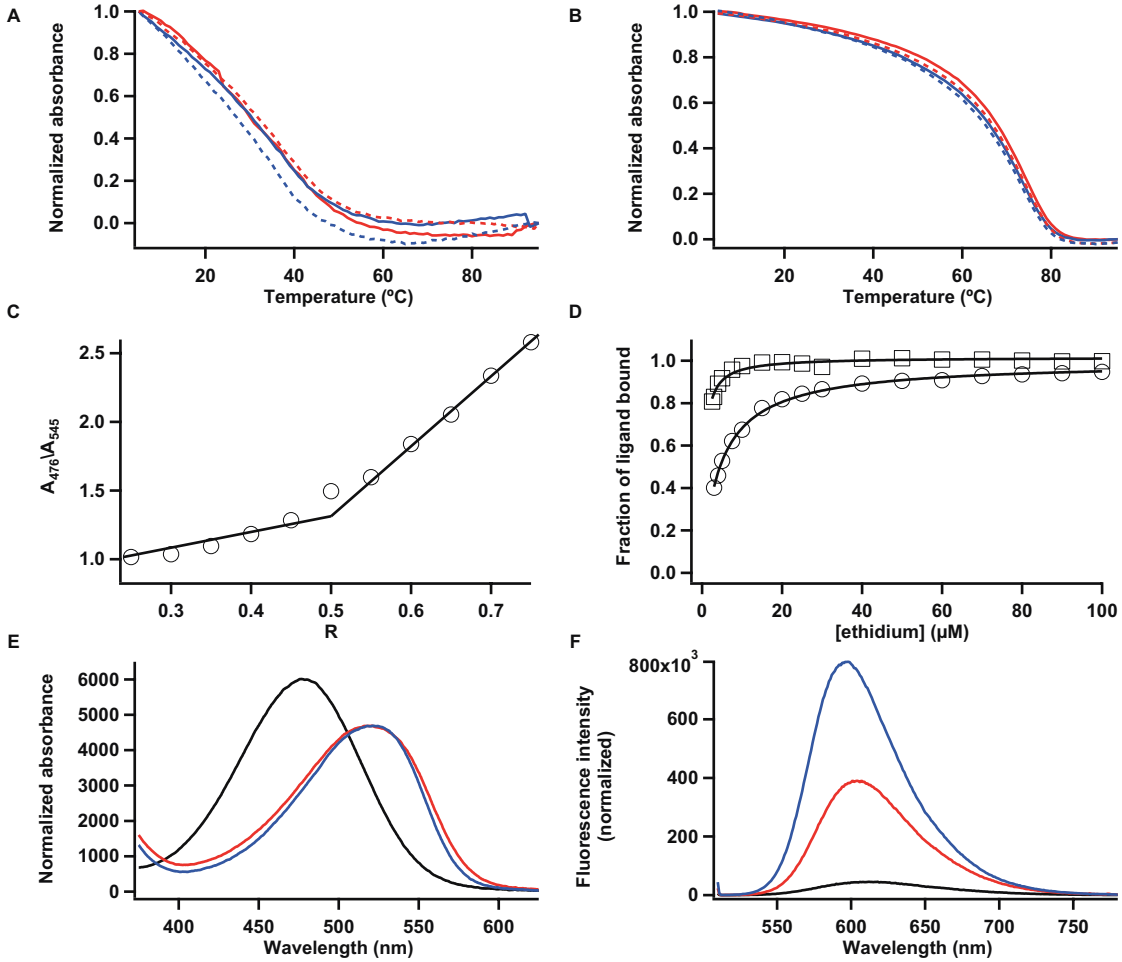

Figure 5.14 UV-Vis thermal denaturation curves for (A) **Pu·Pu** with ethidium and (B) **WC** with ethidium. Red lines represent heating traces, blue lines cooling traces. First heating and cooling traces are solid lines and second heating and cooling traces are dashed lines. (C) Job plot of ethidium with **Pu·Pu** indicating a nearest-neighbor binding stoichiometry. (D) Plots of the fraction of ligand bound to **WC** (open squares) and **Pu·Pu** (open circles) as a function of ligand concentration. (E) UV-Vis spectra of ethidium molecules free (black) and in the presence of **WC** (blue) or **Pu·Pu** (red). (F) Fluorescence emission spectra at 4°C for free ethidium (black), **Pu·Pu** with ethidium (red), and **WC** with ethidium (blue).

Ethidium bound more tightly to **WC** than to **Pu·Pu** by about an order of magnitude; association constants at 4°C were 3.6×10^5 M^{-1} and 4.5×10^5 M^{-1},

respectively (see dilution binding curve in Figure 5.14 D). UV-Vis spectra of ethidium with both types of DNA exhibit the bathochromic (41 nm) and hypochromic shifts associated with an intercalative mode of binding (Figure 5.14 E).

Finally, since ethidium exhibits a fluorescence enhancement upon binding to DNA (see section 5.1.6.3 for discussion), we obtained fluorescence spectra at 4°C for free ethidium, ethidium + **Pu·Pu**, and ethidium + **WC** (Figure 5.14 F). The fluorescence enhancement of the ethidium + **WC** system was approximately 14 times that of free ethidium, while the fluorescence enhancement of the ethidium + **Pu·Pu** system was 7.4 times greater than that of free ethidium. These results suggest similar modes of binding to both types of DNA, resulting in protection from solvent interactions, though this effect is more pronounced for **WC**. This evidence is consistent with the higher ΔT_m observed for ethidium with **WC** and with the higher K_a for ethidium with **WC** that were previously discussed.

5.3.4 Ellipticine Binding to WC and Pu·Pu

Ellipticine was the first small molecule tested that increased the T_m of **Pu·Pu** more than it did the T_m of **WC** (ΔT_m of 24°C versus 10°C). Melt curves were once again superimposable (Figure 5.15 A and B).

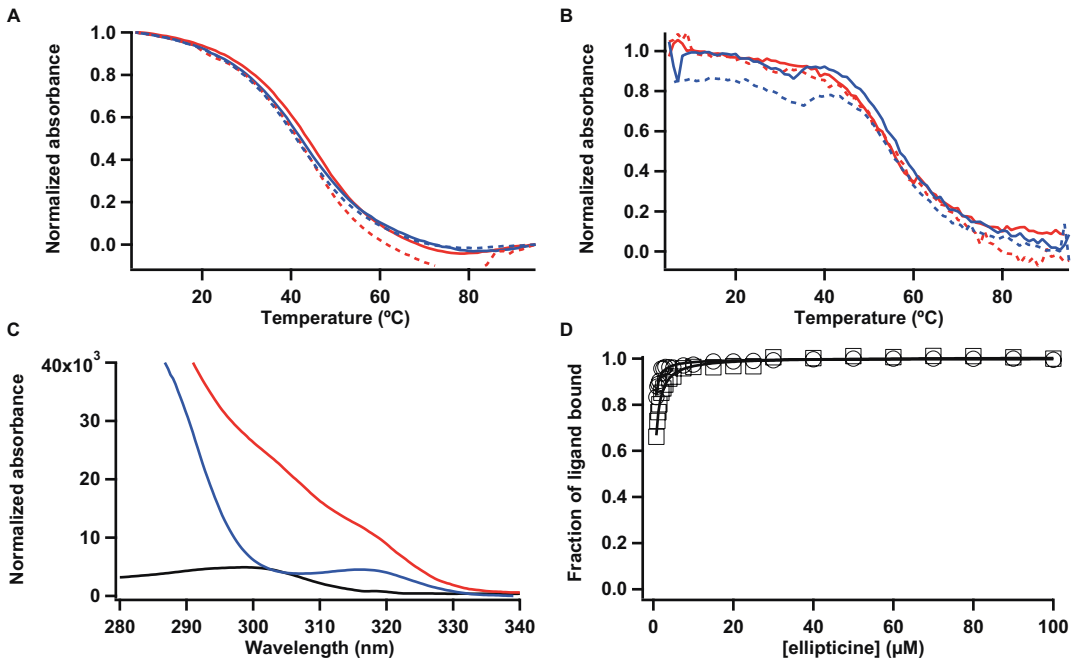

Figure 5.15 UV-Vis thermal denaturation curves for (A) **Pu·Pu** with ellipticine and (B) **WC** with ellipticine. Red lines represent heating traces, blue lines cooling traces. First heating and cooling traces are solid lines and second heating and cooling traces are dashed lines. (C) UV-Vis spectra of ellipticine molecules free (black) and in the presence of **WC** (blue) or **Pu·Pu** (red). (D) Plots of the fraction of ligand bound to **WC** (open squares) and **Pu·Pu** (open circles) as a function of ligand concentration.

Free ellipticine's UV-Vis spectrum contains multiple peaks in the longer-wavelength regions; however, several of these have low extinction coefficients and were not observable under our experimental conditions. An absorption band centered around 300 nm could be monitored in some dilution and thermal denaturation experiments, although **Pu·Pu** also has significant UV-Vis absorbance at 300 nm due to base-pairing tautomer of isoguanine [300-302], overlapping the ellipticine spectrum. We are therefore unable to comment on hypochromicity and red shift for ellipticine and **Pu·Pu** (Figure 5.15 C). For the ellipticine **WC** system, hypochromicity and a bathochromic shift of ~17 nm are observed (Figure 5.15 C).

Ellipticine binds tightly to both types of DNA, with approximately threefold better binding to **Pu·Pu** (15×10^5 M^{-1}) than to **WC (**5.4×10^5 M^{-1}) (Figure 5.15 D). To obtain spectra to generate dilution curves for the overlapping ellipticine + **Pu·Pu** system, a normalized spectrum of drug-free **Pu·Pu** subtracted from each dilution spectrum; the resultant spectra were then analyzed in the usual way. While our ellipticine + **WC** association constant is not as favorable as the literature value (10^7 M^{-1}), the differences may be explained by variations in DNA sequence and/or experimental conditions.

5.3.5 Coralyne Binding to WC and Pu·Pu

Coralyne increases the thermal stability of **Pu·Pu** over **WC** markedly, with ΔT_ms of 25°C and 4°C, respectively. As with other small molecules in this study, two heating and two cooling traces were collected for each DNA + coralyne system; these were reversible and superimposable (Figure 5.16 A and B).

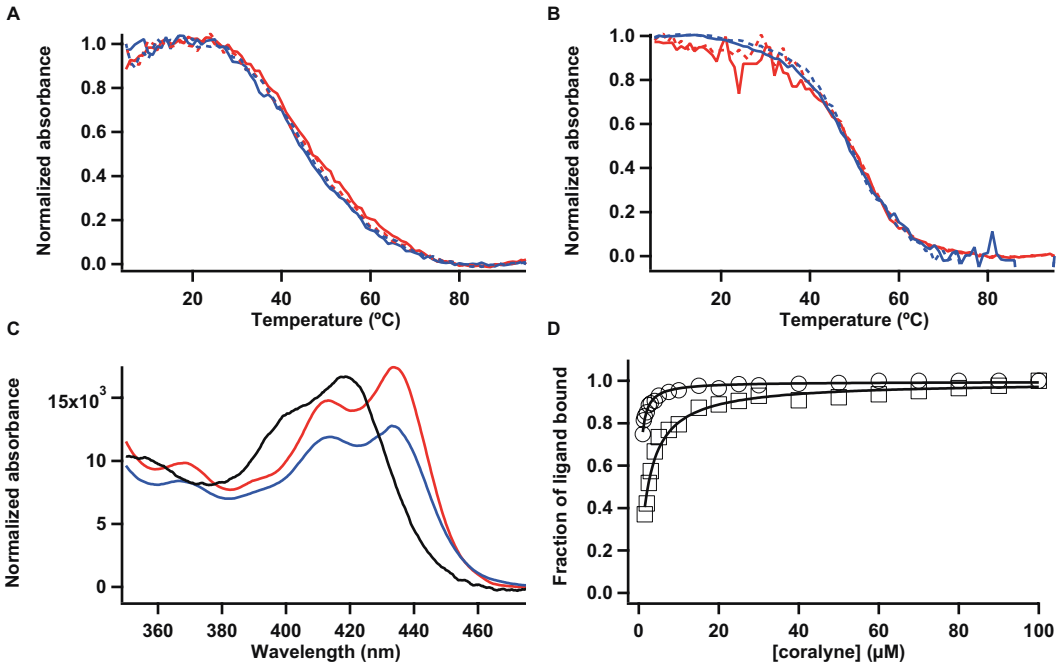

Figure 5.16 UV-Vis thermal denaturation curves for (A) **Pu·Pu** with coralyne and (B) **WC** with coralyne. Red lines represent heating traces, blue lines cooling traces. First heating and cooling traces are solid lines and second heating and cooling traces are dashed lines. (C) UV-Vis spectra of coralyne molecules free (black) and in the presence of **WC** (blue) or **Pu·Pu** (red). (D) Plots of the fraction of ligand bound to **WC** (open squares) and **Pu·Pu** (open circles) as a function of ligand concentration.

Coralyne's long-wavelength absorbance band has a complex shape; however, coralyne in the presence of either type of DNA displays a bathochromic shift of ~15 nm but the hypochromic effect is more pronounced in the case of **WC** DNA. Spectral shape is indicative of a similar, likely intercalative, mode of binding (Figure 5.16 C).

Of all molecules investigated in this study, coralyne exhibits the strongest preference for **Pu·Pu** (6.7×10^5 M^{-1}) versus **WC** (0.9×10^5 M^{-1}) DNA (Figure 5.16 D); this is not surprising considering coralyne's known binding to A·A mismatches (see sections 5.1.4.1 and 5.1.5.4) and its extended ring system and crescent shape, which should match the size and shape of a purine-purine base pair's stacking surface.

5.3.6 Azacyanine3 Binding to WC and Pu·Pu

Aza3, like coralyne and ellipticine, displayed a slight preference for **Pu·Pu** as opposed to **WC** duplexes, resulting in respective ΔT_ms of 20°C and 10°C. All melts with aza3 were reversible (Figure 5.17 A and B).

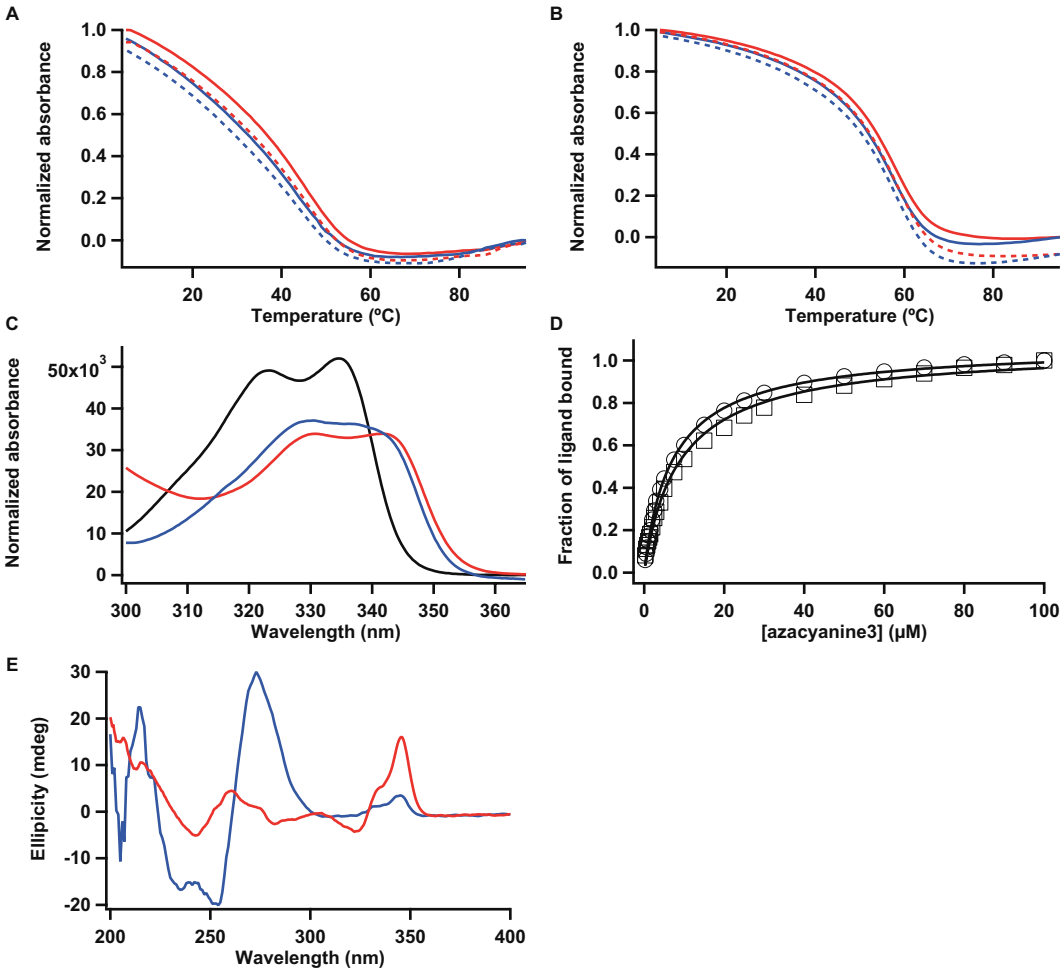

Figure 5.17 UV-Vis thermal denaturation curves for (A) **Pu·Pu** with aza3 and (B) **WC** with aza3. Red lines represent heating traces, blue lines cooling traces. First heating and cooling traces are solid lines and second heating and cooling traces are dashed lines. (C) UV-Vis spectra of aza3 molecules free (black) and in the presence of **WC** (blue) or **Pu·Pu** (red). (D) Plots of the fraction of ligand bound to **WC** (open squares) and **Pu·Pu** (open circles) as a function of ligand concentration. (E) Circular dichroism spectra at 5°C for **Pu·Pu** with aza3 (red) and **WC** with aza3 (blue).

The UV-Vis absorption spectrum of aza3 has a complex shape with two peaks; this spectrum changes upon association with both **WC** and **Pu·Pu** (Figure 5.17 C). However, the spectral changes differ depending on the type of nucleic acid present. For **Pu·Pu**, a hypochromic effect and a bathochromic shift of the more intense, rightmost peak of ~7 nm are observed. These changes are consistent with an intercalative mode of binding. For **WC**, the leftmost of the two peaks becomes slightly greater in magnitude and the valley that exists between the two peaks in the free drug spectrum nearly disappears. A hypochromic effect also exists in this case.

Although the aza3 + **Pu·Pu** system exhibits a high ΔT_m, the binding affinity of this small molecule to both types of DNA is quite similar, with $K_a = 0.28 \times 10^5\ M^{-1}$ for **Pu·Pu** and $K_a = 0.22 \times 10^5\ M^{-1}$ for **WC** (see binding curves in Figure 5.17 D). Aza3 also exhibits induced CD signals with similar shapes at 5°C for both types of DNA, though the intensity of the ICD in the aza3 + **Pu·Pu** system is greater than for aza3 + **WC** in otherwise identical samples (Figure 5.17 E).

5.3.7 Daunomycin Binding to WC and Pu·Pu

Daunomycin at concentrations consistent with the nearest neighbor exclusion principle in complexes with **WC** DNA resulted in a ΔT_m of 24°C (Figure 5.18 A); melts with daunomycin and **Pu·Pu** were neither monophasic nor reversible, and a melting temperature could not be determined for this system.

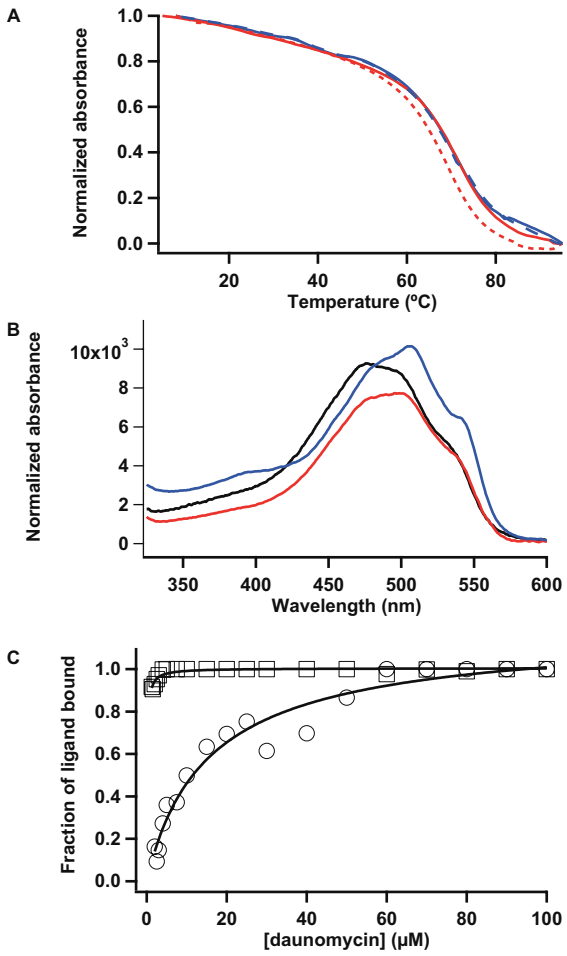

Figure 5.18 (A) UV-Vis thermal denaturation curves for **WC** with daunomycin. Red lines represent heating traces, blue lines cooling traces. First heating and cooling traces are solid lines and second heating and cooling traces are dashed lines. (B) UV-Vis spectra of daunomycin molecules free (black) and in the presence of **WC** (blue) or **Pu·Pu** (red). (C) Plots of the fraction of ligand bound to **WC** (open squares) and **Pu·Pu** (open circles) as a function of ligand concentration.

For daunomycin, a bathochromic shift of ~22 nm is observed for the highest long-wavelength drug peak upon complexing with **Pu·Pu;** a bathochromic shift of ~28 nm is observed with **WC** (Figure 5.18 B). With both types of DNA, we see a similar change in the shape of daunomycin's UV-Vis absorbance spectrum, although this drug has an intricately-shaped spectrum with multiple peaks, and intensities of these peaks relative to one another are somewhat different in the presence of different types of DNA.

Consistent with our hypothesis, daunomycin binds poorly to **Pu·Pu** ($K_a = 0.13 \times 10^5\ M^{-1}$) compared to **WC** ($K_a = 17 \times 10^5\ M^{-1}$); this represents a 132-fold preference for Watson-Crick DNA over homopurine DNA (Figure 5.18 C). This preference is not surprising, considering that the daunomycin/**WC** complex is stabilized by contacts between the daunosamine moiety and the minor groove of B-DNA (see section 5.1.5.6 and references therein) and that the minor groove of **Pu·Pu** is expected to be quite different than that of **WC**.

5.3.8 Actinomycin D Binding to WC and Pu·Pu

Melting temperatures could not be determined for actinomycin D with either type of DNA, as melting curves were not monophasic or reversible, even with extended hold times at each temperature before spectra were recorded. An orange-colored precipitate was observed in the quartz cells at the end of each thermal denaturation experiment attempted with actinomycin D. Müller and Crothers observed that actinomycin binding to DNA distorts the helix such that another drug molecule is unlikely to bind closer than six base pairs from the site of the first drug molecule [288]. Thus, the drug:DNA stoichiometry used for these experiments and the low solubility of actinomycin D in water may have contributed to our inability to determine T_m values for our actinomycin/DNA complexes.

In solution with **Pu·Pu**, the long-wavelength region of the UV-Vis absorbance spectrum of actinomycin D displays a bathochromic shift of ~9 nm, while a somewhat greater bathochromic shift of ~14 nm is seen when actinomycin D binds to **WC** (Figure 5.19 A).

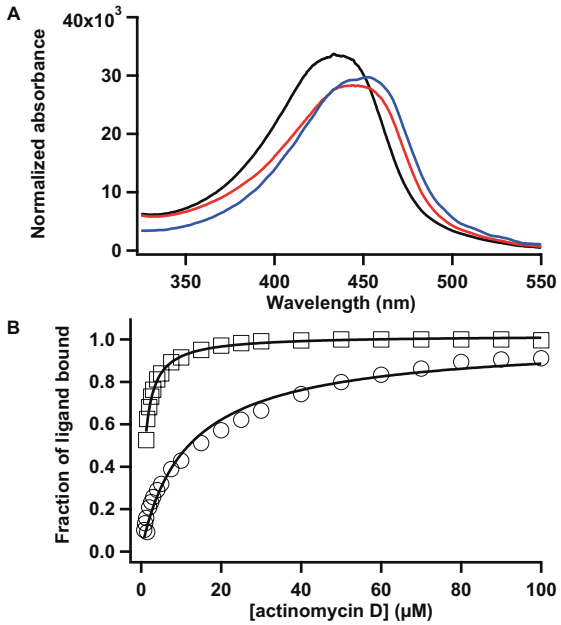

Figure 5.19 (A) UV-Vis spectra of actinomycin D molecules free (black) and in the presence of **WC** (blue) or **Pu·Pu** (red). (B) Plots of the fraction of ligand bound to **WC** (open squares) and **Pu·Pu** (open circles) as a function of ligand concentration.

Actinomycin D's cyclic depsipeptides require approximately 15 minutes to arrive at their equilibrium conformation after mixing at 4°C; other small molecules reached their equilibrium conformations within seconds. Actinomycin D binding to **WC** is about one order of magnitude more favorable ($K_a = 2.0 \times 10^5$ M^{-1}) than to **Pu·Pu** ($K_a = 0.17 \times 10^5$ M^{-1}) (Figure 5.19 B); this is not surprising in light of crystal structures which show the cyclic depsipeptides making specific contacts within the minor groove of B-form DNA (see section 5.1.5.7).

5.3.9 Minor Groove Width

No structural data exists for mixed-sequence antiparallel homopurine DNA duplexes, though polyxanthylic acid can self-pair in a duplex structure and has been crystallized and determined to share some similarities, but not to be identical with, A-form nucleic acids [163]; glycosidic carbons are noted to be 0.2 nm further apart in the

poly(X) structure than in a purine-pyrimidine pair, due to the larger sizes of the base pairs. Denise Okafor in the Hud laboratory performed molecular dynamics simulations in order to determine what helical parameters might be altered in mixed-sequence antiparallel homopurine duplexes such as **Pu·Pu**. Based on our spectroscopic findings with actinomycin D and daunomycin, we anticipated that the minor groove of **Pu·Pu** might be significantly different from that of **WC**; our MD simulations supported this hypothesis (Figures 5.20 and 5.21).

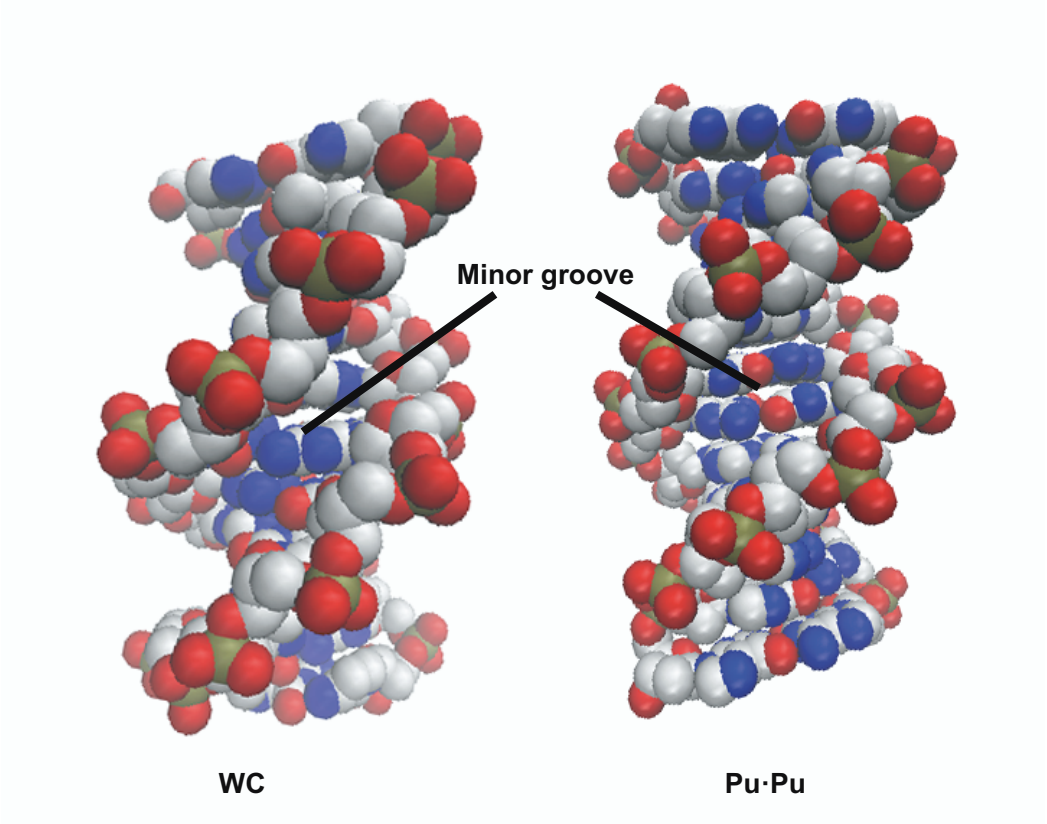

Figure 5.20 Three-dimensional representations of the **WC** and **Pu·Pu** duplexes with the minor groove indicated. Hydrogen atoms have been removed for clarity; carbon atoms are represented in white, phosphorus in gold, oxygen in red, and nitrogen in blue.

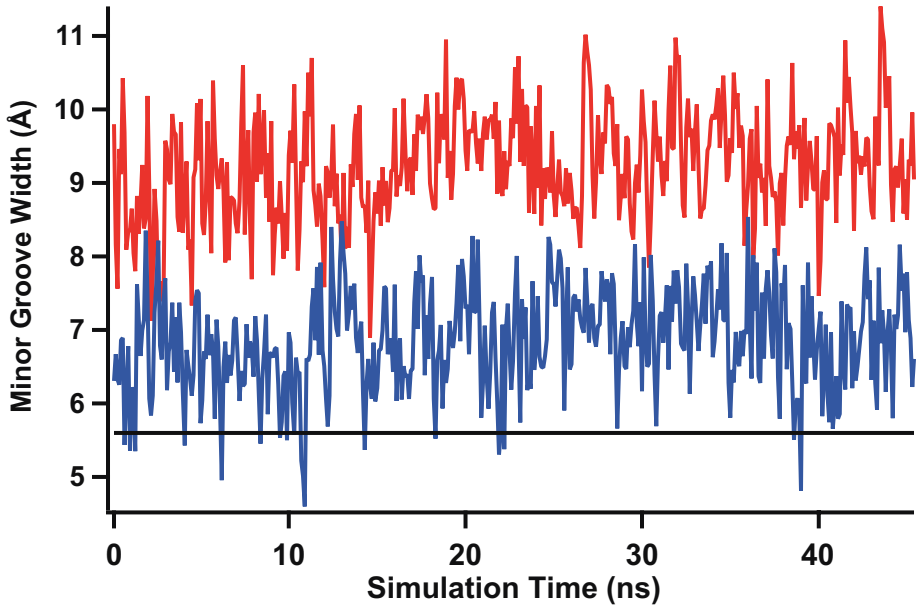

Figure 5.21 Minor groove width of models of the **Pu·Pu** duplex (red) and the **WC** duplex (blue), from 45 ns of MD simulation using the AMBER force field, as calculated using the CURVES program. Black line represents the minor groove width of canonical B-form DNA, as generated by the nucgen program within the AMBER11 software package.

Over the course of the simulations, the width of the minor groove of **Pu·Pu** was about 0.2 nm greater than that of **WC** (Figure 5.21), which is closely-aligned with the poly(X) crystal structure data of Arnott and colleagues.

5.4 DISCUSSION

5.4.1 Binding Mode

Though atomic-resolution structural data is not available, these results of our experiments are consistent with an intercalative mode of binding to **Pu·Pu** for all small molecules investigated. Bathochromic shifts indicative of intercalation were observed for proflavine, ethidium, coralyne, aza3, daunomycin, and actinomycin D in the presence of **Pu·Pu**; while this shift could not be observed with ellipticine due to an overlap in the absorbance spectrum of **Pu·Pu** with that of ellipticine, we assume that this drug is also

intercalating the purine-purine base pairs. Additionally, hypochromicity was observed in the long-wavelength absorbance bands of ethidium, aza3, daunomycin, and actinomycin D in the presence of **Pu·Pu**; this is also consistent with intercalation of these molecules into the homopurine duplex.

Other evidence also supports an intercalative mode of binding. Ethidium is protected from fluorescence quenching by solvent when bound to **Pu·Pu**; this is also true of ethidium molecules which intercalate natural DNA. Job plots of both proflavine and ethidium with **Pu·Pu** indicate a binding stoichiometry of one small molecule for every two base pairs, consistent with the nearest neighbor exclusion principle.

Since it is likely, based on the available evidence, that these molecules intercalate **Pu·Pu**, they are useful as models for the putative molecular midwives which may have aided the assembly of the earliest nucleic acids.

5.4.2 Size- and Shape-Mediated Binding

As expected, molecules which are known to bind strongly to Watson-Crick DNA interacted more favorably with **WC** than with **Pu·Pu**; these include proflavine and ethidium, which selectively bound **WC** over **Pu·Pu** by factors of 2.4 and 7.7, respectively (Table 5.1). These molecules increased the T_m of the **WC** duplex by more substantial amounts than they did for **Pu·Pu**. Proflavine and ethidium have smaller surface areas available for stacking than many of the other molecules investigated in this study; these areas are similar in size to purine-pyrimidine base pairs.

On the other hand, molecules with larger stacking surface areas and crescent shapes, such as aza3, coralyne, and ellipticine, exhibited selectivity towards the **Pu·Pu** duplex and its larger base pairs. This was not unexpected, given that these molecules are

known to bind to other polypurine structures [157, 286]. Coralyne preferentially bound **Pu·Pu** by a factor of 7.4 (Table 5.1); this was the most favorable interaction with **Pu·Pu** of all drugs investigated. Ellipticine's presence in this group is perhaps somewhat surprising, given that it is one of the tightest known binders to natural DNA [281]; however, a closer inspection of its structure (Section 5.1.5.3) reveals an extended ring system and a crescent shape – properties it shares with coralyne and aza3.

Table 5.1 Changes in melting temperature, association constants, and association constant ratios for small molecules with **Pu·Pu** and **WC**. $^{*}T_m$ rather than ΔT_m. †Melting experiment not completed due to aggregation of small molecule.

Ligand	ΔT_m (°C)		K_a ($\times 10^5$ M^{-1})		K_a(**Pu·Pu**)/K_a(**WC**)
	Pu·Pu	**WC**	**Pu·Pu**	**WC**	
None	20*	46*	--	--	--
Proflavine	9	20	1.4	3.3	$(2.4)^{-1}$
Ethidium	11	25	0.45	3.6	$(7.7)^{-1}$
Ellipticine	24	10	15	5.4	2.8
Coralyne	25	4	6.7	0.9	7.4
Aza3	20	10	0.28	0.22	1.3
Daunomycin	N/A†	24	0.13	17	$(132)^{-1}$
Actinomycin D	N/A†	N/A†	0.17	2.0	$(12)^{-1}$

5.4.3 The Role of the Minor Groove

MD simulations demonstrate that the minor groove of **Pu·Pu** is 0.2 nm wider than that of **WC**. It is therefore not surprising that molecules such as actinomycin D and daunomycin, which are known to make specific contacts in the minor groove of DNA [185, 186], exhibit 12-fold and 132-fold preferences for **WC**. The difference in groove widths between the two duplexes may also partially explain ethidium's preference for **WC**, as ethidium's pendant ring also projects into the minor groove [187].

5.5 CONCLUDING REMARKS

We have demonstrated that crescent-shaped molecules with extended ring systems bind preferentially to an antiparallel homopurine duplex over a standard Watson-Crick duplex. Further, we have shown that natural products such as actinomycin D and daunomycin, which possess moieties that project into the minor groove, selectively associate with a mixed-sequence purine-pyrimidine duplex. These molecules, which are produced by bacteria, would have evolved after the advent of life and would have been selected to target existing nucleic acid structures. It should, in theory, be possible to design molecules with moieties that preferentially bind to the minor groove of homopurine DNA, as well.

One aspect of molecular recognition of nucleic acids by small molecules that has not been investigated in this work is thermodynamics. The contributions of enthalpy and entropy to the drug-binding process are not obvious. It has previously been demonstrated that merely switching a backbone linkage from 3′-5′ to 2′-5′ alters the free energy of intercalation from being an entropically-driven event to an entropically-disfavored and enthalpically-driven one [261]. It is thus not possible to predict the driving force behind intercalation of an all-purine duplex; further calorimetric studies could provide valuable insights on this process.

In this work, we have roughly equated increased T_m in the presence of small molecules with preferential stabilization by said molecules of either **WC** or **Pu·Pu**. While it is common in the literature to speak of increased thermal stability, in reality, this practice results in the oversimplification of a complex process [303]. We expect that the association constants – and thus the free energies of association – for the small molecules

with each type of DNA will vary with temperature, though in this work these values were only determined at 4°C. Further, in the presence of a ligand, ΔT_m also depends on the number of binding sites and cooperativity of the binding sites [303]. In our experiments, we assumed adherence to the nearest neighbor exclusion principle; however, some of the molecules studied (e.g. actinomycin D) may bind to nucleic acid duplexes at lower stoichiometries [304] or may also interact favorably with single stranded structures, especially **Pu·Pu** with its greater surface area available for stacking.

An inherent difficulty in working with **Pu·Pu** is its low melting temperature. Increasing the number of hydrogen bonds between base pairs by incorporating the diaminopurine·xanthine pair (see Figure 1.19) should result in a duplex with a higher T_m; experiments with a xanthine analog have yielded T_ms for all-purine DNA comparable to those of standard DNA duplexes [43]. An equivalent series of experiments with the diaminopurine·xanthine, diaminopurine·inosine, and adenine·xanthine base pairs could shed additional light on the role of the third hydrogen bond, as well as on the wider applicability of the effects of minor groove substituents. Uncompensated amines in the minor groove have been demonstrated to be destabilizing in purine-purine pairs inserted into purine-pyrimidine duplexes [305], though electronegative atoms such as oxygen provide additional stability when present in the minor grooves of purine-pyrimidine duplexes [306].

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

6.1 FORMATION OF PURINE NUCLEOBASES AND MOLECULAR RECOGNITION OF PURINE-PURINE BASE PAIRS

We have observed the facile synthesis of several purine bases from formamide, including purine, hypoxanthine, guanine, adenine, 2,6-diaminopurine, and 6,8-diaminopurine. Through modifications of the reaction protocol, such as the addition of small amounts of water, the use of different wavelengths of UV and visible light for irradiation, the use of different minerals as catalysts, the spiking of small molecule intermediates into reactions, and the improvement of analytical techniques, it may be possible to detect other purine nucleobases, including xanthine and isoguanine. A mechanism of assembly, such as stacking with a molecular midwife, would be an important step in the process of forming a proto-RNA.

Glycosylation of nucleobases and backbone formation would still be required. Since ribose is such a minor product of the formose reaction [133], it is conceivable that the first nucleic acids contained sugars other than ribose. Since the phosphodiester linkage is high in energy, the first nucleosides may have been held together by reversible linkages [188] such as acetals [152] or imines formed from aldehydes and amines [307]. Such linkages should allow for thermodynamically-favored products with the correct base pairing to form, even in the absence of polymerase enzymes.

Over time, we can envision the evolution of pre-RNA nucleic acid sequences which contain the purine bases synthesized in the formamide reaction as recognition

elements. Though the hypothesized homopurine nucleic acid sequences must have later given way to more thermally-stable mixed purine-pyrimidine sequences, all-purine sequences might be easier to form initially due to favorable stacking interactions [145].

6.2 FURTHER EXPERIMENTS IN FORMAMIDE CHEMISTRY

6.2.1 Mineral Reactions

Saladino and Di Mauro have extensively cataloged the products of formamide reactions conducted in the presence of a variety of mineral catalysts [36-38, 79]. However, their experiments were thermal only, not photochemical; further, all of their reactions were conducted under an air atmosphere. It would be of great interest to choose additional prebiotically-relevant minerals, especially those with some photocatalytic properties, and conduct formamide reactions catalyzed by these minerals in the presence of different wavelengths of UV and visible light and under different atmospheres.

A sealed reaction center would be required for precise manipulation of atmospheric gas partial pressures, or the use of gas mixtures. A tunable light source would allow the study of the contributions of different wavelengths of light to formamide chemistry. Since it has been demonstrated that adding water to formamide still allows for the production of nucleobases after extended periods of heating at low temperatures [40], a reactor that is capable of simulating day-night wet-dry cycles might also be of interest.

6.2.2 Synthesis of Larger Molecules in Formamide Reaction Mixtures

Numerous purines, pyrimidines, and triazines have been identified as formamide reaction products [36, 40]. Extended ring systems have also been detected in the HCN polymer [115] (see also Section 1.8). It is reasonable to assume that condensation products having three or more rings are generated in formamide reactions; these are

likely to be minor products which are present at concentrations lower than the detection limits of the instruments used thus far to analyze formamide reaction mixtures. Larger ring systems are attractive as potential intercalators or stacking surfaces for the assembly of early nucleic acid analogs [188]; a one-pot synthesis of both nucleobases and "molecular midwives" would be particularly appealing.

6.2.2.1 Preliminary Results

Preliminary experiments indicate the synthesis of large heterocycles (three or more rings) from formamide. A diverse variety of larger molecules appear to be present in these reactions; this is not surprising considering the complexity of chromatograms for one- and two-ring compounds.

6.2.2.1.1 Reaction Conditions

All reagents were purchased from VWR or Sigma Aldrich and used as received. 5 g of sodium pyrophosphate was placed in a 250 mL round-bottom flask with 100 mL of formamide. The flask was equipped with a water condenser and was heated in a sand bath at 160°C, with stirring, for 48 hours.

6.2.2.1.2 Sample Work-Up

The reaction flask was removed from the heat source and allowed to cool to room temperature. It was diluted to 1 L with nanoPure water and filtered through 0.8 μ syringe filters (Corning). If the diluted reaction mixture was allowed to settle, either before or after filtration, large assemblies formed in the solution; these could be disturbed by shaking the flask in which the solution was stored, but would form again if the mixture was allowed to settle.

After filtration, the reaction mixture was loaded onto C18 Sep Pak solid phase extraction cartridges (Waters). It was hypothesized that large, hydrophobic products would favorably associate with the C18 stationary phase while more soluble products would wash through with the loading solution. This would allow us to concentrate minor products from the bulk solution. A dark, tarry substance was collected on the cartridges. Retained products were eluted with various solvents; methanol and isopropyl alcohol removed some components with UV absorbance, and dimethyl sulfoxide (DMSO) caused more compounds to elute. The Sep Pak cartridges, however, remained dark-colored; not all retained compounds could be eluted.

Various means were explored to purify and isolate reaction products with the goal of further characterization by crystallization, mass spectrometry, UV-Vis spectrophotometry, and/or NMR spectroscopy. HPLC purification was tested on several stationary phases including C8 (Agilent Zorbax), amino (Phenomenex), and C18 (Phenomenex); reinjection of fractions from these tests indicated the presence of multiple compounds. An analytical injection using a Kinetex column with a PFP (pentafluorophenyl) stationary phase (Phenomenex) appeared to provide the best separation.

Methanol and isopropyl alcohol could be removed by rotary evaporation or with a stream of nitrogen prior to further analysis of samples eluted with these solvents. Additionally, methanol and isopropyl alcohol are acceptable solvents for injection into an LC or LC-MS instrument. However, DMSO has a boiling point of 189°C and could not easily be evaporated. We attempted to remove DMSO by loading samples eluted with this solvent onto a cellulose phosphate cation exchange column (Whatman) and flowing

increasing concentrations of sodium phosphate buffer solution (pH = 7) over the column. While this method did successfully remove DMSO from samples, the phosphate buffer interfered with analysis by mass spectrometry.

Preparatory TLC (thin layer chromatography) of fractions eluted from the Sep Pak cartridges with DMSO appears promising; this method separates DMSO from products of interest chromatographically, without the need for the introduction of salts incompatible with mass spectrometry. Preparatory TLC appears to perform some separation of the sample, as well. If repeated several times, the preparatory TLC process could generate enough of the various reaction products to carry through for further analysis including attempts to crystallize, determination of pK_a, or even characterization of their DNA-binding properties.

6.2.2.1.3 UV-Vis and Mass Spectrometry

At several stages of the purification process, samples were subjected to mass spectral and UV-Vis analysis. Several high-mass compounds were observed in a fraction of formamide reaction eluted from a Sep Pak with methanol; these had mass-to-charge ratios exceeding 500 Da (Figure 6.1). Considering that many nitrogen heterocycles are observed as products of formamide chemistry [36], and considering the assemblies observed (see Section 6.2.2.1.2), it is conceivable that these molecules could possess large aromatic stacking surfaces similar to those associated with porphyrins.

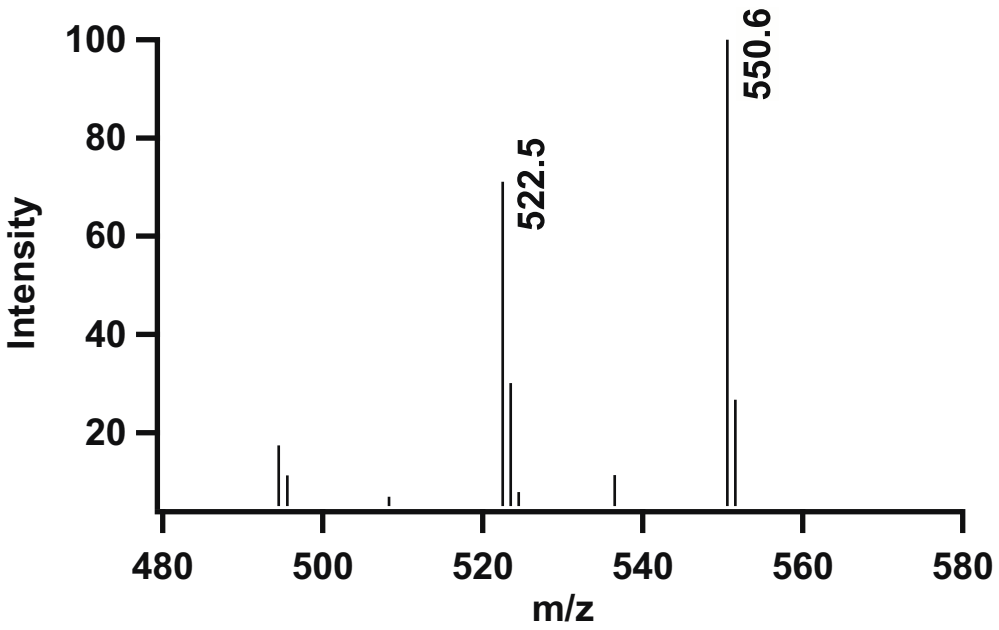

Figure 6.1 Ions with m/z from 480 Da to 580 Da and their relative abundances; from a direct mass spec injection of formamide reaction fractions eluted from a C18 Sep Pak cartridge with methanol. Selected masses are marked. This mass spectrum is the result of a direct injection of one fraction on an Agilent 6100 series single quadrupole mass spectrometer.

Further, fractions from the cellulose phosphate cation exchange column (see Section 6.2.2.1.2) were analyzed with a UV-Vis spectrophotometer (Figure 6.2). This analysis revealed the presence of several compounds with wavelengths of maximum absorption at 300 nm and above. Purine and pyrimidine nucleobases generally absorb strongest at 260 nm, while larger, multicyclic molecules such as DNA intercalators have longer wavelengths of maximum absorption (see Section 5.2.2.2).

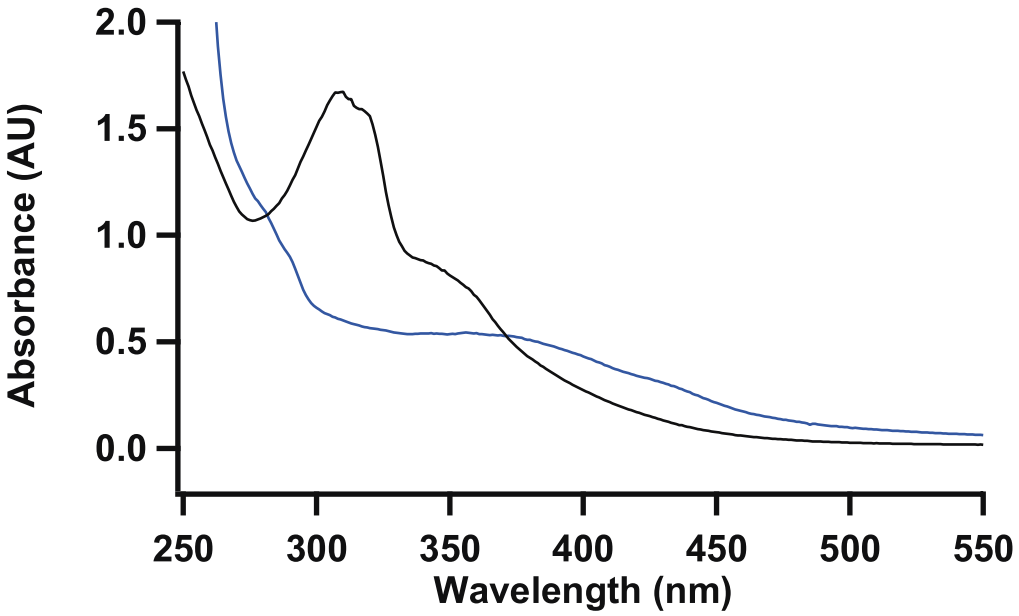

Figure 6.2 UV-Vis spectra of two fractions with long-wavelength absorbance from formamide reaction fractions eluted first from a C18 Sep Pak cartridge with DMSO and then from a cellulose phosphate cation exchange column with increasing concentrations of phosphate buffer.

Coupled with the mass spectrometry data showing compounds with high masses, the UV-Vis spectra showing long-wavelength absorbance strongly support the synthesis of larger molecules in formamide reactions.

6.2.2.2 Future Directions

Attempts to further purify and/or isolate large heterocycles from formamide reactions have thus far been unsuccessful. Normal phase liquid chromatography with hydrophobic solvents could aid HPLC separations of reaction products; further, reverse phase methods employing the PFP stationary phase should also be tested.

Thus far, only one set of reaction conditions has been explored for the synthesis of large heterocycles from formamide (sodium pyrophosphate catalyst, 160°C). The use

of different catalysts, different atmospheres, and the addition of UV irradiation to the system should also be explored, once a purification method has been optimized.

6.3 CONNECTION TO THE PRE-RNA WORLD

The first nucleic acids and proto-RNAs must have been easy to assemble from molecules that were available on the prebiotic Earth, whether these compounds were synthesized on the planet or delivered exogenously via meteorites, comets, or IDPs. These polymers would almost certainly have been heterogeneous in nature, and any function they possessed would have been inferior to contemporary RNA or DNA. Even assuming that sequence information is transferred with a low mutation rate, and that errors in sequence can be corrected if backbone linkages are reversible, different sugars and backbone linkages would likely lead to sub-optimal folding, which would negatively affect any potentially catalytic activities these sequences were capable of carrying out.

However, given time, proto-RNAs which evolved to incorporate bases, sugars, and linkages with superior functionalities would have been able to outcompete their heterogeneous predecessors, eventually leading to the RNA World and later to the life of today, governed by the Central Dogma of molecular biology and incorporating proteins and peptides along with RNA and DNA.

REFERENCES

[1] J. van Wyhe, *The complete work of charles darwin online* 2012, http://darwin-online.org.uk/ accessed May 3, 2012.

[2] J. W. Schopf, A. B. Kudryavtsev, D. G. Agresti, T. J. Wdowiak and A. D. Czaja, *Laser-raman imagery of earth's earliest fossils* Nature 2002, *416*(6876): p. 73-76.

[3] D. Wacey, M. R. Kilburn, M. Saunders, J. Cliff and M. D. Brasier, *Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of western australia* Nature Geoscience 2011, *4*(10): p. 698-702.

[4] N. Glansdorff, Y. Xu and B. Labedan, *The last universal common ancestor: Emergence, constitution and genetic legacy of an elusive forerunner* Biology Direct 2008, 3(29).

[5] H. Martin, F. Albarede, P. Claeys, M. Gargaud, B. Marty, A. Morbidelli and D. L. Pinti, *Building of a habitable planet* Earth Moon and Planets 2006, *98*(1-4): p. 97-151.

[6] U. J. Meierhenrich, *Comment life in its uniqueness remains difficult to define in scientific terms* Journal of Biomolecular Structure & Dynamics 2012, *29*(4): p. 641-642.

[7] F. H. C. Crick, *Central dogma of molecular biology* Nature 1970, *227*(5258): p. 561-563.

[8] S. W. Fox, *The evolutionary significance of phase separated microsystems* Origins of Life 1976, *7*(1): p. 49-68.

[9] S. W. Fox, *Metabolic microspheres - origins and evolution* Naturwissenschaften 1980, *67*(8): p. 378-383.

[10] S. W. Fox, *Synthesis of life in the lab - defining a protoliving system* Quarterly Review of Biology 1991, *66*(2): p. 181-185.

[11] P. A. Bachmann, P. Walde, P. L. Luisi and J. Lang, *Self-replicating reverse micelles and chemical autopoiesis* Journal of the American Chemical Society 1990, *112*(22): p. 8200-8201.

[12] P. A. Bachmann, P. Walde, P. L. Luisi and J. Lang, *Self-replicating micelles aqueous micelles and enzymatically driven reactions in reverse micelles* Journal of the American Chemical Society 1991, *113*(22): p. 8204-8209.

[13] S. E. Maurer, D. W. Deamer, J. M. Boncella and P. A. Monnard, *Chemical evolution of amphiphiles: Glycerol monoacyl derivatives stabilize plausible prebiotic membranes* Astrobiology 2009, *9*(10): p. 979-987.

[14] P. A. Monnard and D. W. Deamer, *Membrane self-assembly processes: Steps toward the first cellular life* Anatomical Record 2002, *268*(3): p. 196-207.

[15] A. Pohorille and D. Deamer, *Self-assembly and function of primitive cell membranes* Research in Microbiology 2009, *160*(7): p. 449-456.

[16] V. N. Kompanichenko, *Stages of transition from precellular organic microsystems to primary prokaryotic communities* Biology Bulletin 2011, *38*(5): p. 542-550.

[17] E. Chargaff, R. Lipshitz and C. Green, *Composition of the desoxypentose nucleic acids of 4 genera of sea-urchin* Journal of Biological Chemistry 1952, *195*(1): p. 155-160.

[18] D. Elson and E. Chargaff, *On the desoxyribonucleic acid content of sea urchin gametes* Experientia 1952, *8*(4): p. 143-145.

[19] S. Zamenhof, G. Brawerman and E. Chargaff, *On the desoxypentose nucleic acids from several microorganisms* Biochimica Et Biophysica Acta 1952, *9*(4): p. 402-405.

[20] J. D. Watson and F. H. C. Crick, *Genetical implications of the structure of deoxyribonucleic acid* Nature 1953, *171*(4361): p. 964-967.

[21] J. D. Watson and F. H. C. Crick, *Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid* Nature 1953, *171*(4356): p. 737-738.

[22] M. Meselson and F. W. Stahl, *The replication of DNA in escherichia coli* Proceedings of the National Academy of Sciences 1958, *44*(7): p. 671-682.

[23] A. Rich and D. R. Davies, *A new 2 stranded helical structure - polyadenylic acid and polyuridylic acid* Journal of the American Chemical Society 1956, *78*(14): p. 3548-3549.

[24] F. H. C. Crick, *Origin of genetic code* Journal of Molecular Biology 1968, *38*(3): p. 367-379.

[25] L. E. Orgel, Evolution of the genetic apparatus - a review *Cold Spring Harbor Symposia on Quantitative Biology* 1987, 52: p. 9-16.

[26] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling and T. R. Cech, *Self-splicing RNA - auto-excision and auto-cyclization of the ribosomal-RNA intervening sequence of tetrahymena* Cell 1982, *31*(1): p. 147-157.

[27] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace and S. Altman, *The RNA moiety of ribonuclease-p is the catalytic subunit of the enzyme* Cell 1983, *35*(3): p. 849-857.

[28] W. Gilbert, *Origin of life - the RNA world* Nature 1986, *319*(6055): p. 618.

[29] E. K. Y. Leung, N. Suslov, N. Tuttle, R. Sengupta and J. A. Piccirilli in *The mechanism of peptidyl transfer catalysis by the ribosome*, *Vol. 80* Eds.: R. D. Kornberg, C. R. H. Raetz, J. E. Rothman and J. W. Thorner), Annual Reviews, Palo Alto, 2011, pp. 527-555.

[30] P. Nissen, J. Hansen, N. Ban, P. B. Moore and T. A. Steitz, *The structural basis of ribosome activity in peptide bond synthesis* Science 2000, *289*(5481): p. 920-930.

[31] P. B. Moore and T. A. Steitz, *After the ribosome structures: How does peptidyl transferase work?* Rna-a Publication of the Rna Society 2003, *9*(2): p. 155-159.

[32] P. B. Moore and T. A. Steitz, *The ribosome revealed* Trends in Biochemical Sciences 2005, *30*(6): p. 281-283.

[33] S. A. Benner, H.-J. Kim and Z. Yang, Setting the stage: The history, chemistry, and geobiology behind RNA *Cold Spring Harbor Perspectives in Biology* 2012, 4: p. 1-13.

[34] M. P. Robertson and G. F. Joyce, The origins of the RNA world *Cold Spring Harbor Perspectives in Biology* 2010, p. 1-22.

[35] T. R. Cech, The RNA worlds in context *Cold Spring Harbor Perspectives in Biology* 2011, p. 1-5.

[36] R. Saladino, C. Crestini, F. Ciciriello, G. Costanzo and E. Di Mauro, *Formamide chemistry and the origin of informational polymers* Chemistry & Biodiversity 2007, *4*(4): p. 694-720.

[37] R. Saladino, V. Neri, C. Crestini, G. Costanzo, M. Graciotti and E. Di Mauro, *Synthesis and degradation of nucleic acid components by formamide and iron sulfur minerals* Journal of the American Chemical Society 2008, *130*(46): p. 15512-15518.

[38] R. Saladino, V. Neri, C. Crestini, G. Costanzo, M. Graciotti and E. Di Mauro, *The role of the formamide/zirconia system in the synthesis of nucleobases and biogenic carboxylic acid derivatives* Journal of Molecular Evolution 2010, *71*(2): p. 100-110.

[39] J. Oro, *Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive earth conditions* Nature 1961, *191*(479): p. 1193-1194.

[40] H. L. Barks, R. Buckley, G. A. Grieves, E. Di Mauro, N. V. Hud and T. M. Orlando, *Guanine, adenine, and hypoxanthine production in uv-irradiated formamide solutions: Relaxation of the requirements for prebiotic purine nucleobase formation* Chembiochem 2010, *11*(9): p. 1240-1243.

[41] S. A. Benner, A. Ricardo and M. A. Carrigan, *Is there a common chemical model for life in the universe?* Current Opinion in Chemical Biology 2004, *8*(6): p. 672-689.

[42] S. A. Benner, *Understanding nucleic acids using synthetic chemistry* Accounts of Chemical Research 2004, *37*(10): p. 784-797.

[43] B. D. Heuberger and C. Switzer, *An alternative nucleobase code: Characterization of purine-purine DNA double helices bearing guanine-isoguanine and diaminopurine 7-deaza-xanthine base pairs* Chembiochem 2008, *9*(17): p. 2779-2783.

[44] T. R. Battersby, M. Albalos and M. J. Friesenhahn, *An unusual mode of DNA duplex association: Watson-crick interaction of all-purine deoxyribonucleic acids* Chemistry & Biology 2007, *14*(5): p. 525-531.

[45] G. Wachtershauser, *An all-purine precursor of nucleic-acids* Proceedings of the National Academy of Sciences of the United States of America 1988, *85*(4): p. 1134-1135.

[46] A. N. Simonov, O. P. Pesturlova, L. G. Matvienko, V. N. Snytnikov, O. A. Snytnikova, Y. P. Tsentalovich and V. N. Parmon, *Possible prebiotic synthesis of monosaccharides from formaldehyde in presence of phosphates* Advances in Space Research 2007, *40*(11): p. 1634-1640.

[47] A. L. Weber, *The sugar model: Catalysis by amines and amino acid products* Origins of Life and Evolution of the Biosphere 2001, *31*(1-2): p. 71-86.

[48] R. Breslow, *On the mechanism of the formose reaction* Tetrahedron Letters 1959, *1*(21): p. 22-26.

[49] S. G. Trevino, N. Zhang, M. P. Elenko, A. Luptak and J. W. Szostak, *Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity* Proceedings of the National Academy of Sciences of the United States of America 2011, *108*(33): p. 13492-13497.

[50] S. L. Miller, J. W. Schopf and A. Lazcano, *Oparin's "origin of life": Sixty years later* Journal of Molecular Evolution 1997, *44*(4): p. 351-353.

[51] I. Fry, *The origins of research into the origins of life* Endeavour 2006, *30*(1): p. 24-28.

[52] J. D. Bernal, *The origin of life*, The World Publishing Company, Cleveland, OH, 1967.

[53] J. Haldane in *The origin of life*, *Vol.* (Ed. J. D. Bernal), The World Publishing Company, Cleveland, OH, 1967, pp. 242-249.

[54] A. I. Oparin in *The origin of life*, *Vol.* (Ed. J. D. Bernal), The World Publishing Company, Cleveland, OH, 1967, pp. 199-241.

[55] S. L. Miller, *A production of amino acids under possible primitive earth conditions* Science 1953, *117*(3046): p. 528-529.

[56] S. L. Miller and H. C. Urey, *Organic compound synthesis on the primitive earth* Science 1959, *130*(3370): p. 245-251.

[57] J. Farquhar, A. L. Zerkle and A. Bekker, *Geological constraints on the origin of oxygenic photosynthesis* Photosynthesis Research 2011, *107*(1): p. 11-36.

[58] G. Ramstein, *Climates of the earth and cryosphere evolution* Surveys in Geophysics 2011, *32*(4-5): p. 329-350.

[59] N. H. Sleep, D. K. Bird and E. C. Pope, *Serpentinite and the dawn of life* Philosophical Transactions of the Royal Society B-Biological Sciences 2011, *366*(1580): p. 2857-2869.

[60] M. J. Russell, *The hazy details of early earth's atmosphere* Science 2010, *330*(6005): p. 754-754.

[61] E. T. Wolf and O. B. Toon, *The hazy details of early earth's atmosphere response* Science 2010, *330*(6005): p. 754-755.

[62] E. T. Wolf and O. B. Toon, *Fractal organic hazes provided an ultraviolet shield for early earth* Science 2010, *328*(5983): p. 1266-1268.

[63] F. Tian, O. B. Toon, A. A. Pavlov and H. De Sterck, *A hydrogen-rich early earth atmosphere* Science 2005, *308*(5724): p. 1014-1017.

[64] S. Miyakawa, H. Yamanashi, K. Kobayashi, H. J. Cleaves and S. L. Miller, *Prebiotic synthesis from co atmospheres: Implications for the origins of life* Proceedings of the National Academy of Sciences of the United States of America 2002, *99*(23): p. 14628-14631.

[65] L. V. Berkner and L. C. Marshall, *On origin and rise of oxygen concentration in earths atmosphere* Journal of the Atmospheric Sciences 1965, *22*(3): p. 225-261.

[66] J. F. Kasting, K. J. Zahnle, J. P. Pinto and A. T. Young, *Sulfur, ultraviolet-radiation, and the early evolution of life* Origins of Life and Evolution of the Biosphere 1989, *19*(2): p. 95-108.

[67] C. Sagan, *On origin and planetary distribution of life* Radiation Research 1961, *15*(2): p. 174-192.

[68] C. Sagan, *Ultraviolet selection pressure on earliest organisms* Journal of Theoretical Biology 1973, *39*(1): p. 195-200.

[69] H. J. Cleaves and S. L. Miller, *Oceanic protection of prebiotic organic compounds from uv radiation* Proceedings of the National Academy of Sciences of the United States of America 1998, *95*(13): p. 7260-7263.

[70] F. Gomez, A. Aguilera and R. Amils, *Soluble ferric iron as an effective protective agent against uv radiation: Implications for early life* Icarus 2007, *191*(1): p. 352-359.

[71] S. D. Senanayake and H. Idriss, *Photocatalysis and the origin of life: Synthesis of nucleoside bases from formamide on tio2(001) single surfaces* Proceedings of the National Academy of Sciences of the United States of America 2006, *103*(5): p. 1194-1198.

[72] R. Saladino, U. Ciambecchini, C. Crestini, G. Costanzo, R. Negri and E. Di Mauro, *One-pot tio2-catalyzed synthesis of nucleic bases and acyclonucleosides from formamide: Implications for the origin of life* Chembiochem 2003, *4*(6): p. 514-521.

[73] V. M. Canuto, J. S. Levine, T. R. Augustsson, C. L. Imhoff and M. S. Giampapa, *The young sun and the atmosphere and photochemistry of the early earth* Nature 1983, *305*(5932): p. 281-286.

[74] J. F. Kasting, J. B. Pollack and D. Crisp, *Effects of high co2 levels on surface-temperature and atmospheric oxidation-state of the early earth* Journal of Atmospheric Chemistry 1984, *1*(4): p. 403-428.

[75] J. P. Ferris and C. T. Chen, *Chemical evolution .26. Photochemistry of methane, nitrogen, and water mixtures as a model for atmosphere of primitive earth* Journal of the American Chemical Society 1975, *97*(11): p. 2962-2967.

[76] J. S. Hubbard, J. P. Hardy and N. H. Horowitz, *Photocatalytic production of organic compounds from co and h2o in a simulated martian atmosphere* Proceedings of

the National Academy of Sciences of the United States of America 1971, *68*(3): p. 574-&.

[77] J. P. Pinto, G. R. Gladstone and Y. L. Yung, *Photochemical production of formaldehyde in earths primitive atmosphere* Science 1980, *210*(4466): p. 183-184.

[78] R. Saladino, C. Crestini, U. Ciambecchini, F. Ciciriello, G. Costanzo and E. Di Mauro, *Synthesis and degradation of nucleobases and nucleic acids by formamide in the presence of montmorillonites* Chembiochem 2004, *5*(11): p. 1558-1566.

[79] R. Saladino, V. Neri and C. Crestini, *Role of clays in the prebiotic synthesis of sugar derivatives from formamide* Philosophical Magazine 2010, *90*(17-18): p. 2329-2337.

[80] U. Shanker, B. Bhushan, G. Bhattacharjee and Kamaluddin, *Formation of nucleobases from formamide in the presence of iron oxides: Implication in chemical evolution and origin of life* Astrobiology 2011, *11*(3): p. 225-233.

[81] G. Costanzo, R. Saladino, C. Crestini, F. Ciciriello and E. Di Mauro, *Nucleoside phosphorylation by phosphate minerals* Journal of Biological Chemistry 2007, *282*(23): p. 16729-16735.

[82] R. Saladino, C. Crestini, V. Neri, F. Ciciriello, G. Costanzo and E. Mauro, *Origin of informational polymers: The concurrent roles of formamide and phosphates* Chembiochem 2006, *7*(11): p. 1707-1714.

[83] A. M. Schoffstall, *Prebiotic phosphorylation of nucleosides in formamide* Origins of Life and Evolution of the Biosphere 1976, *7*(4): p. 399-412.

[84] A. M. Schoffstall, R. J. Barto and D. L. Ramos, *Nucleoside and deoxynucleoside phosphorylation in formamide solutions* Origins of Life and Evolution of the Biosphere 1982, *12*(2): p. 143-151.

[85] J. P. Ferris, *Mineral catalysis and prebiotic synthesis: Montmorillonite-catalyzed formation of RNA* Elements 2005, *1*(3): p. 145-149.

[86] J. P. Ferris, A. R. Hill, R. H. Liu and L. E. Orgel, *Synthesis of long prebiotic oligomers on mineral surfaces* Nature 1996, *381*(6577): p. 59-61.

[87] R. M. Hazen, D. Papineau, W. B. Leeker, R. T. Downs, J. M. Ferry, T. J. McCoy, D. A. Sverjensky and H. X. Yang, *Mineral evolution* American Mineralogist 2008, *93*(11-12): p. 1693-1720.

[88] G. Wachtershauser, *Before enzymes and templates - theory of surface metabolism* Microbiological Reviews 1988, *52*(4): p. 452-484.

[89] G. Wachtershauser, *Pyrite formation, the 1st energy-source for life - a hypothesis* Systematic and Applied Microbiology 1988, *10*(3): p. 207-210.

[90] G. Wachtershauser, *Evolution of the 1st metabolic cycles* Proceedings of the National Academy of Sciences of the United States of America 1990, *87*(1): p. 200-204.

[91] G. Wachtershauser, *Groundworks for an evolutionary biochemistry - the iron sulfur world* Progress in Biophysics & Molecular Biology 1992, *58*(2): p. 85-201.

[92] A. Michalkova, Y. Kholod, D. Kosenkov, L. Gorb and J. Leszczynski, *Viability of pyrite pulled metabolism in the 'iron-sulfur world' theory: Quantum chemical assessment* Geochimica Et Cosmochimica Acta 2011, *75*(7): p. 1933-1941.

[93] M. A. Pasek and D. S. Lauretta, *Aqueous corrosion of phosphide minerals from iron meteorites: A highly reactive source of prebiotic phosphorus on the surface of the early earth* Astrobiology 2005, *5*(4): p. 515-535.

[94] J. C. G. Walker and P. Brimblecombe, *Iron and sulfur in the pre-biologic ocean* Precambrian Research 1985, *28*(3-4): p. 205-222.

[95] R. Osterberg, *On the prebiotic role of iron and sulfur* Origins of Life and Evolution of Biospheres 1997, *27*(5): p. 481-484.

[96] H. D. Holland, *Oceans - possible source of iron in iron-formations* Economic Geology 1973, *68*(7): p. 1169-1172.

[97] M. I. Budyko, *Effect of solar radiation variations on climate of earth* Tellus 1969, *21*(5): p. 611-619.

[98] A. S. Endal and K. H. Schatten, *The faint young sun-climate paradox - continental influences* Journal of Geophysical Research-Oceans and Atmospheres 1982, *87*(C9): p. 7295-7302.

[99] K. H. Schatten and A. S. Endal, *The faint young sun - climate paradox - volcanic influences* Geophysical Research Letters 1982, *9*(12): p. 1309-1311.

[100] C. Sagan and G. Mullen, *Earth and mars - evolution of atmospheres and surface temperatures* Science 1972, *177*(4043): p. 52-56.

[101] S. Miyakawa, H. J. Cleaves and S. L. Miller, *The cold origin of life: A. Implications based on the hydrolytic stabilities of hydrogen cyanide and formamide* Origins of Life and Evolution of the Biosphere 2002, *32*(3): p. 195-208.

[102] H. J. Cleaves, K. E. Nelson and S. L. Miller, *The prebiotic synthesis of pyrimidines in frozen solution* Naturwissenschaften 2006, *93*(5): p. 228-231.

[103] S. Miyakawa, H. J. Cleaves and S. L. Miller, *The cold origin of life: B. Implications based on pyrimidines and purines produced from frozen ammonium cyanide solutions* Origins of Life and Evolution of the Biosphere 2002, *32*(3): p. 209-218.

[104] C. Menor-Salvan, D. M. Ruiz-Bermejo, M. I. Guzman, S. Osuna-Esteban and S. Veintemillas-Verdaguer, *Synthesis of pyrimidines and triazines in ice: Implications for the prebiotic chemistry of nucleobases* Chemistry-a European Journal 2009, *15*(17): p. 4411-4418.

[105] J. L. Bada, C. Bigham and S. L. Miller, *Impact melting of frozen oceans on the early earth - implications for the origin of life* Proceedings of the National Academy of Sciences of the United States of America 1994, *91*(4): p. 1248-1250.

[106] M. T. Rosing, D. K. Bird, N. H. Sleep and C. J. Bjerrum, *No climate paradox under the faint early sun* Nature 2010, *464*(7289): p. 744-747.

[107] J. Oro and A. P. Kimball, *Synthesis of purines under possible primitive earth conditions .1. Adenine from hydrogen cyanide* Archives of Biochemistry and Biophysics 1961, *94*(2): p. 217-227.

[108] D. P. Cruikshank, H. Imanaka and C. M. D. Ore in *Tholins as coloring agents on outer solar system bodies*, *Vol. 36* Eds.: M. Bernstein, R. NavarroGonzalez and R. Raulin), Elsevier Science Ltd, Oxford, 2005, pp. 178-183.

[109] J. H. Debes, A. J. Weinberger and G. Schneider, *Complex organic materials in the circumstellar disk of hr 4796a* Astrophysical Journal Letters 2008, *673*(2): p. L191-L194.

[110] B. N. Khare, C. Sagan, H. Ogino, B. Nagy, C. Er, K. H. Schram and E. T. Arakawa, *Amino-acids derived from titan tholins* Icarus 1986, *68*(1): p. 176-184.

[111] G. D. McDonald, W. R. Thompson, M. Heinrich, B. N. Khare and C. Sagan, *Chemical investigation of titan and triton tholins* Icarus 1994, *108*(1): p. 137-145.

[112] C. Sagan and B. N. Khare, *Tholins - organic-chemistry of inter-stellar grains and gas* Nature 1979, *277*(5692): p. 102-107.

[113] I. Mamajanov and J. Herzfeld, *Hcn polymers characterized by ssnmr: Solid state reaction of crystalline tetramer (diaminomaleonitrile)* Journal of Chemical Physics 2009, 130(13).

[114] I. Mamajanov and J. Herzfeld, *Hcn polymers characterized by solid state nmr: Chains and sheets formed in the neat liquid* Journal of Chemical Physics 2009, 130(13).

[115] M. Ruiz-Bermejo, J. L. de la Fuente, C. Rogero, C. Menor-Salván, S. Osuna-Esteban and J. A. Martín-Gago, *New insights into the characterization of 'insoluble black hcn polymers'* Chemistry & Biodiversity 2012, *9*(1): p. 25-40.

[116] A. Hill and L. E. Orgel, *Synthesis of adenine from hcn tetramer and ammonium formate* Origins of Life and Evolution of the Biosphere 2002, *32*(2): p. 99-102.

[117] J. M. Notley and M. Spiro, *Purification of formamide and rate of its reaction with dissolved water* Journal of the Chemical Society B-Physical Organic 1966, 4): p. 362-366.

[118] V. Campos, A. C. G. Marigliano and H. N. Solimo, *Density, viscosity, refractive index, excess molar volume, viscosity, and refractive index deviations and their correlations for the (formamide plus water) system. Isobaric (vapor plus liquid) equilibrium at 2.5 kpa* Journal of Chemical and Engineering Data 2008, *53*(1): p. 211-216.

[119] M. P. Callahan, K. E. Smith, H. J. Cleaves, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House and J. P. Dworkin, *Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases* Proceedings of the National Academy of Sciences of the United States of America 2011, *108*(34): p. 13995-13998.

[120] J. P. Ferris and L. E. Orgel, *Aminomalononitrile and 4-amino-5-cyanoimidazole in hydrogen cyanide polymerization and adenine synthesis* Journal of the American Chemical Society 1965, *87*(21): p. 4976-4977.

[121] J. P. Ferris and L. E. Orgel, *An unusual photochemical rearrangement in synthesis of adenine from hydrogen cyanide* Journal of the American Chemical Society 1966, *88*(5): p. 1074.

[122] J. P. Ferris and L. E. Orgel, *Studies in prebiotic synthesis .I. Aminomalononitrile and 4-amino-5-cyanoimidazole* Journal of the American Chemical Society 1966, *88*(16): p. 3829-3831.

[123] R. A. Sanchez, J. P. Ferris and L. E. Orgel, *Studies in prebiotic synthesis .2. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide* Journal of Molecular Biology 1967, *30*(2): p. 223-253.

[124] R. A. Sanchez, J. P. Ferris and L. E. Orgel, *Studies in prebiotic synthesis .4. Conversion of 4-aminoimidazole-5-carbonitrile derivatives to purines* Journal of Molecular Biology 1968, *38*(1): p. 121-128.

[125] K. T. Suzuki, H. Yamada and M. Hirobe, *Thermal fission and reformation of c-n bond in formamide as studied by c-13-n-15 coupling* Journal of the Chemical Society-Chemical Communications 1978, 11): p. 485-486.

[126] H. Yamada and T. Okamoto, *One-step synthesis of purine ring from formamide* Chemical & Pharmaceutical Bulletin 1972, *20*(3): p. 623-624.

[127] R. Saladino, C. Crestini, F. Ciciriello, S. Pino, G. Costanzo and E. Di Mauro, *From formamide to RNA: The roles of formamide and water in the evolution of chemical information* Research in Microbiology 2009, *160*(7): p. 441-448.

[128] R. Saladino, C. Crestini, C. Cossetti, E. Di Mauro and D. Deamer, *Catalytic effects of murchison material: Prebiotic synthesis and degradation of RNA precursors* Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life 2011, *41*(5): p. 437-451.

[129] R. Saladino, C. Crestini, V. Neri, J. R. Brucato, L. Colangeli, F. Ciciriello, E. Di Mauro and G. Costanzo, *Synthesis and degradation of nucleic acid components by formamide and cosmic dust analogues* Chembiochem 2005, *6*(8): p. 1368-1374.

[130] A. P. Johnson, H. J. Cleaves, J. P. Dworkin, D. P. Glavin, A. Lazcano and J. L. Bada, *The miller volcanic spark discharge experiment* Science 2008, *322*(5900): p. 404-404.

[131] A. Butlerov, *Formation of monosaccharides from formaldehyde* Comptes rendus de l'Académie des sciences 1861, *53*(p. 145-147.

[132] I. V. Delidovich, O. P. Taran, A. N. Simonov, L. G. Matvienko and V. N. Parmon, *Photoinduced catalytic synthesis of biologically important metabolites from formaldehyde and ammonia under plausible "Prebiotic" Conditions* Advances in Space Research 2011, *48*(3): p. 441-449.

[133] R. Shapiro, *Prebiotic ribose synthesis: A critical analysis* Origins of Life and Evolution of Biospheres 1988, *18*(1): p. 71-85.

[134] G. Zubay and T. Mui, *Prebiotic synthesis of nucleotides* Origins of Life and Evolution of the Biosphere 2001, *31*(1-2): p. 87-102.

[135] J. P. Ferris, O. S. Zamek, A. M. Altbuch and H. Freiman, *Chemical evolution .18. Synthesis of pyrimidines from guanidine and cyanoacetaldehyde* Journal of Molecular Evolution 1974, *3*(4): p. 301-309.

[136] M. P. Robertson and S. L. Miller, *An efficient prebiotic synthesis of cytosine and uracil* Nature 1995, *375*(6534): p. 772-774.

[137] A. W. Schwartz and G. J. F. Chittenden, *Synthesis of uracil and thymine under simulated prebiotic conditions* Biosystems 1977, *9*(2-3): p. 87-92.

[138] G. Cooper, N. Kimmich, W. Belisle, J. Sarinana, K. Brabham and L. Garrel, *Carbonaceous meteorites as a source of sugar-related organic compounds for the early earth* Nature 2001, *414*(6866): p. 879-883.

[139] O. Botta and J. L. Bada, *Extraterrestrial organic compounds in meteorites* Surveys in Geophysics 2002, *23*(5): p. 411-467.

[140] C. Chyba and C. Sagan, *Endogenous production, exogenous delivery and impact-shock synthesis of organic-molecules - an inventory for the origins of life* Nature 1992, *355*(6356): p. 125-132.

[141] W. D. Fuller, L. E. Orgel and R. A. Sanchez, *Studies in prebiotic synthesis .7. Solid-state synthesis of purine nucleosides* Journal of Molecular Evolution 1972, *1*(3): p. 249-257.

[142] H. D. Bean, Y. H. Sheng, J. P. Collins, F. A. L. Anet, J. Leszczynski and N. V. Hud, *Formation of a beta-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction* Journal of the American Chemical Society 2007, *129*(31): p. 9556-9557.

[143] A. W. Schwartz, *Prebiotic phosphorylation-nucleotide synthesis with apatite* Biochimica Et Biophysica Acta 1972, *281*(4): p. 477-480.

[144] M. W. Powner, B. Gerland and J. D. Sutherland, *Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions* Nature 2009, *459*(7244): p. 239-242.

[145] P. O. P. Ts'o, I. S. Melvin and A. C. Olson, *Interaction and association of bases and nucleosides in aqueous solutions* Journal of the American Chemical Society 1963, *85*(9): p. 1289-1296.

[146] J. C. Chaput, J. K. Ichida and J. W. Szostak, *DNA polymerase-mediated DNA synthesis on a tna template* Journal of the American Chemical Society 2003, *125*(4): p. 856-857.

[147] A. Eschenmoser, *Searching for nucleic acid alternatives* Chimia 2005, *59*(11): p. 836-850.

[148] L. L. Zhang, A. Peritz and E. Meggers, *A simple glycol nucleic acid* Journal of the American Chemical Society 2005, *127*(12): p. 4174-4175.

[149] K. C. Schneider and S. A. Benner, *Oligonucleotides containing flexible nucleoside analogs* Journal of the American Chemical Society 1990, *112*(1): p. 453-455.

[150] P. A. Giannaris and M. J. Damha, *Oligoribonucleotides containing 2',5'-phosphodiester linkages exhibit binding selectivity for 3',5'-RNA over 3',5'-ssdna* Nucleic Acids Research 1993, *21*(20): p. 4742-4749.

[151] P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm and B. Norden, *DNA-like double helix formed by peptide nucleic-acid* Nature 1994, *368*(6471): p. 561-563.

[152] H. D. Bean, F. A. L. Anet, I. R. Gould and N. V. Hud, *Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA* Origins of Life and Evolution of the Biosphere 2006, *36*(1): p. 39-63.

[153] J. P. Ferris and G. Ertem, *Oligomerization of ribonucleotides on montmorillonite - reaction of the 5'-phosphorimidazolide of adenosine* Science 1992, *257*(5075): p. 1387-1389.

[154] E. D. Horowitz, A. E. Engelhart, M. C. Chen, K. A. Quarles, M. W. Smith, D. G. Lynn and N. V. Hud, *Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world* Proceedings of the National Academy of Sciences of the United States of America 2010, *107*(12): p. 5288-5293.

[155] A. Rich, *Formation of 2-stranded and 3-stranded helical molecules by polyinosinic acid and polyadenylic acid* Nature 1958, *181*(4608): p. 521-525.

[156] A. Rich, F. H. C. Crick, J. D. Watson and D. R. Davies, *Molecular structure of polyadenylic acid* Journal of Molecular Biology 1961, *3*(1): p. 71-86.

[157] O. P. Cetinkol and N. V. Hud, *Molecular recognition of poly(a) by small ligands: An alternative method of analysis reveals nanomolar, cooperative and shape-selective binding* Nucleic Acids Research 2009, *37*(2): p. 611-621.

[158] I. S. Joung, O. P. Cetinkol, N. V. Hud and E. Thomas, *Molecular dynamics simulations and coupled nucleotide substitution experiments indicate the nature of a center dot a base pairing and a putative structure of the coralyne-induced homo-adenine duplex* Nucleic Acids Research 2009, *37*(22): p. 7715-7727.

[159] J. T. Davis, *G-quartets 40 years later: From 5 '-gmp to molecular biology and supramolecular chemistry* Angewandte Chemie-International Edition 2004, *43*(6): p. 668-698.

[160] S. Arnott and P. J. Bond, *Triple-stranded polynucleotide helix containing only purine bases* Science 1973, *181*(4094): p. 68-69.

[161] M. Fikus and D. Shugar, *Properties of poly-xanthylic acid and its reactions with potentially complementary homopolynucleotides* Acta Biochimica Polonica 1969, *16*(1): p. 55-82.

[162] J. C. Chaput and C. Switzer, *A DNA pentaplex incorporating nucleobase quintets* Proceedings of the National Academy of Sciences of the United States of America 1999, *96*(19): p. 10614-10619.

[163] S. Arnott, R. Chandrasekaran, A. W. Day, L. C. Puigjaner and L. Watts, *Double-helical structures for polyxanthylic acid* Journal of Molecular Biology 1981, *149*(3): p. 489-505.

[164] S. Uesugi, Y. Oda, M. Ikehara, Y. Kawase and E. Ohtsuka, *Identification of i-a mismatch base-pairing structure in DNA* Journal of Biological Chemistry 1987, *262*(15): p. 6965-6968.

[165] K. Groebke, J. Hunziker, W. Fraser, L. Peng, U. Diederichsen, K. Zimmermann, A. Holzner, C. Leumann and A. Eschenmoser, *Why pentose- and not hexose-nucleic acids? Purine-purine pairing in homo-DNA: Guanine, isoguanine, 2,6-diaminopurine, and xanthine* Helvetica Chimica Acta 1998, *81*(3): p. 375-474.

[166] C. L. Defty and J. R. Marsden, *Melphalan in regional chemotherapy for locally recurrent metastatic melanoma* Current Topics in Medicinal Chemistry 2012, *12*(1): p. 53-60.

[167] M. Christmann, B. Verbeek, W. P. Roos and B. Kaina, *O(6)-methylguanine-DNA methyltransferase (mgmt) in normal tissues and tumors: Enzyme activity, promoter methylation and immunohistochemistry* Biochimica Et Biophysica Acta-Reviews on Cancer 2011, *1816*(2): p. 179-190.

[168] T. Iwamoto, Y. Hiraku, S. Oikawa, H. Mizutani, M. Kojima and S. Kawanishi, *DNA intrastrand cross-link at the 5'-ga-3' sequence formed by busulfan and its role in the cytotoxic effect* Cancer Science 2004, *95*(5): p. 454-458.

[169] R. A. Alderden, M. D. Hall and T. W. Hambley, *The discovery and development of cisplatin* Journal of Chemical Education 2006, *83*(5): p. 728.

[170] S. Walker, R. Landovitz, W. D. Ding, G. A. Ellestad and D. Kahne, *Cleavage behavior of calicheamicin gamma 1 and calicheamicin t* Proceedings of the National Academy of Sciences 1992, *89*(10): p. 4608-4612.

[171] C. Carvalho, R. X. Santos, S. Cardoso, S. Correia, P. J. Oliveira, M. S. Santos and P. I. Moreira, *Doxorubicin: The good, the bad and the ugly effect* Current Medicinal Chemistry 2009, *16*(25): p. 3267-3285.

[172] A. A. Zuma, D. P. Cavalcanti, M. C. P. Maia, W. de Souza and M. C. M. Motta, *Effect of topoisomerase inhibitors and DNA-binding drugs on the cell proliferation and ultrastructure of trypanosoma cruzi* International Journal of Antimicrobial Agents 2011, *37*(5): p. 449-456.

[173] C. Bailly and J. B. Chaires, *Sequence-specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues* Bioconjugate Chemistry 1998, *9*(5): p. 513-538.

[174] R. D. Anderson and N. A. Berger, *Mutagenicity and carcinogenicity of topoisomerase-interactive agents* Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis 1994, *309*(1): p. 109-142.

[175] P. R. Turner and W. A. Denny, *The mutagenic properties of DNA minor-groove binding ligands* Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis 1996, *355*(1-2): p. 141-169.

[176] J. B. Chaires, *A thermodynamic signature for drug,äìdna binding mode* Archives of Biochemistry and Biophysics 2006, *453*(1): p. 26-31.

[177] W. Saenger, *Principles of nucleic acid structure*, Springer-Verlag, New York, 1984.

[178] H. M. Sobell, C. Tsai, S. C. Jain and S. G. Gilbert, *Visualization of drug-nucleic acid interactions at atomic resolution .3. Unifying structural concepts in understanding drug-DNA interactions and their broader implications in understanding protein-DNA interactions* Journal of Molecular Biology 1977, *114*(3): p. 333-365.

[179] T. Maehigashi, O. Persil, N. V. Hud and L. D. Williams, *Crystal structure of proflavine in complex with a DNA hexamer duplex* Protein Data Bank, 2009.

[180] H. Ihmels and D. Otto in *Intercalation of organic dye molecules into double-stranded DNA general principles and recent developments*, *Vol. 258* (Ed. F. Wurthner), Springer-Verlag Berlin, Berlin, 2005, pp. 161-204.

[181] S. N. Rao and P. A. Kollman, *Molecular mechanical simulations on double intercalation of 9-amino acridine into d(cgcgcgc).D(gcgcgcg) - analysis of the physical basis for the neighbor-exclusion principle* Proceedings of the National Academy of Sciences of the United States of America 1987, *84*(16): p. 5735-5739.

[182] R. D. Blake, *Informational biopolymers of genes and gene expression*, University Science Books, Sausalito, California, 2005.

[183] T. E. Haran and U. Mohanty, *The unique structure of a-tracts and intrinsic DNA bending* Quarterly Reviews of Biophysics 2009, *42*(01): p. 41-81.

[184] M. C. Vega, I. GarcÍA SÁEz, J. AymamÍ, R. Eritja, G. A. Van Der Marel, J. H. Van Boom, A. Rich and M. Coll, *Three-dimensional crystal structure of the a-tract DNA dodecamer d(cgcaaatttgcg) complexed with the minor-groove-binding drug hoechst 33258* European Journal of Biochemistry 1994, *222*(3): p. 721-726.

[185] S. Kamitori and F. Takusagawa, *Crystal-structure of the 2/1 complex between d(gaagcttc) and the anticancer drug actinomycin-d* Journal of Molecular Biology 1992, *225*(2): p. 445-456.

[186] M. H. Moore, W. N. Hunter, B. L. d'Estaintot and O. Kennard, *DNA-drug interactions: The crystal structure of d(cgatcg) complexed with daunomycin* Journal of Molecular Biology 1989, *206*(4): p. 693-705.

[187] C. C. Tsai, S. C. Jain and H. M. Sobell, *X-ray crystallographic visualization of drug nucleic acid intercalative binding - structure of an ethidium-dinucleoside monophosphate crystalline complex, ethidium - 5-iodouridylyl(3'-5')adenosine* Proceedings of the National Academy of Sciences of the United States of America 1975, *72*(2): p. 628-632.

[188] N. V. Hud and F. A. L. Anet, *Intercalation-mediated synthesis and replication: A new approach to the origin of life* Journal of Theoretical Biology 2000, *205*(4): p. 543-562.

[189] M. P. Robertson, M. Levy and S. L. Miller, *Prebiotic synthesis of diaminopyrimidine and thiocytosine* Journal of Molecular Evolution 1996, *43*(6): p. 543-550.

[190] A. Eschenmoser and E. Loewenthal, *Chemistry of potentially prebiological natural-products* Chemical Society Reviews 1992, *21*(1): p. 1-16.

[191] E. C. Taylor, H. M. Loux, E. A. Falco and G. H. Hitchings, *Pyrimidopteridines by oxidative self-condensation of aminopyrimidines* Journal of the American Chemical Society 1955, *77*(8): p. 2243-2248.

[192] A. Eschenmoser, *Etiology of potentially primordial biomolecular structures: From vitamin b12 to the nucleic acids and an inquiry into the chemistry of life's origin: A retrospective* Angewandte Chemie-International Edition 2011, *50*(52): p. 12412-12472.

[193] A. I. Oparin, *Coacervate drops as models of pre biological systems*, Elsevier Publishing Co., Inc., New York, N.Y., U.S.A 1971.

[194] G. Wachtershauser, *Life in a ligand sphere* Proceedings of the National Academy of Sciences of the United States of America 1994, *91*(10): p. 4283-4287.

[195] A. Y. Mulkidjanian, A. Y. Bychkov, D. V. Dibrova, M. Y. Galperin and E. V. Koonin, *Origin of first cells at terrestrial, anoxic geothermal fields* Proceedings of the National Academy of Sciences 2012, *109*(14): p. E821-E830.

[196] R. Saladino, C. Crestini, G. Costanzo, R. Negri and E. Di Mauro, *A possible prebiotic synthesis of purine, adenine, cytosine, and 4(3h)-pyrimidinone from formamide: Implications for the origin of life* Bioorganic & Medicinal Chemistry 2001, *9*(5): p. 1249-1253.

[197] C. S. Cockell, *The ultraviolet history of the terrestrial planets - implications for biological evolution* Planetary and Space Science 2000, *48*(2-3): p. 203-214.

[198] S. N. Vinh, H. L. Abbott, M. M. Dawley, T. M. Orlando, J. Leszczynski and T. N. Minh, *Theoretical study of formamide decomposition pathways* Journal of Physical Chemistry A 2011, *115*(5): p. 841-851.

[199] F. Duvernay, A. Trivella, F. Borget, S. Coussan, J. P. Aycard and T. Chiavassa, *Matrix isolation fourier transform infrared study of photodecomposition of formimidic acid* Journal of Physical Chemistry A 2005, *109*(49): p. 11155-11162.

[200] E. B. Watson and T. M. Harrison, *Zircon thermometer reveals minimum melting conditions on earliest earth* Science 2005, *308*(5723): p. 841-844.

[201] T. M. Harrison, J. Blichert-Toft, W. M√ºller, F. Albarede, P. Holden and S. J. Mojzsis, *Response to comment on "Heterogeneous hadean hafnium: Evidence of continental crust at 4.4 to 4.5 ga"* Science 2006, *312*(5777): p. 1139.

[202] J. M. Gingell, N. J. Mason, H. Zhao, I. C. Walker and M. R. F. Siggel, *Vuv optical-absorption and electron-energy-loss spectroscopy of formamide* Chemical Physics 1997, *220*(1-2): p. 191-205.

[203] J. P. Ferris, P. C. Joshi, E. H. Edelson and J. G. Lawless, *Chemical evolution .30. Hcn - plausible source of purines, pyrimidines and amino-acids on primitive earth* Journal of Molecular Evolution 1978, *11*(4): p. 293-311.

[204] D. Bockelee-Morvan, D. C. Lis, J. E. Wink, D. Despois, J. Crovisier, R. Bachiller, D. J. Benford, N. Biver, P. Colom, J. K. Davies, E. Gerard, B. Germain, M. Houde, D. Mehringer, R. Moreno, G. Paubert, T. G. Phillips and H. Rauer, *New molecules found in comet c/1995 o1 (hale-bopp) - investigating the link between cometary and interstellar material* Astronomy & Astrophysics 2000, *353*(3): p. 1101-1114.

[205] M. J. Drake, *Origin of water in the terrestrial planets* Meteoritics & Planetary Science 2005, *40*(4): p. 519-527.

[206] D. T. Halfen, V. Ilyushin and L. M. Ziurys, *Formation of peptide bonds in space: A comprehensive study of formamide and acetamide in sgr b2(n)* Astrophysical Journal 2011, *743*(1): p. 1-12.

[207] A. Ricardo, M. A. Carrigan, A. N. Olcott and S. A. Benner, *Borate minerals stabilize ribose* Science 2004, *303*(5655): p. 196-196.

[208] N. Lahav, *Minerals and the origin of life - hypotheses and experiments in heterogeneous chemistry* Heterogeneous Chemistry Reviews 1994, *1*(2): p. 159-179.

[209] J. W. Morgan and E. Anders, *Chemical composition of earth, venus, and mercury* Proceedings of the National Academy of Sciences 1980, *77*(12): p. 6973-6977.

[210] A. Bekker, J. F. Slack, N. Planavsky, B. Krapez, A. Hofmann, K. O. Konhauser and O. J. Rouxel, *Iron formation: The sedimentary product of a complex interplay among mantle, tectonic, oceanic, and biospheric processes* Economic Geology 2010, *105*(3): p. 467-508.

[211] P. S. Braterman, A. G. Cairnssmith and R. W. Sloper, *Photooxidation of hydrated fe-2+ - significance for banded iron formations* Nature 1983, *303*(5913): p. 163-164.

[212] A. P. Chandra and A. R. Gerson, *Pyrite (fes2) oxidation: A sub-micron synchrotron investigation of the initial steps* Geochimica Et Cosmochimica Acta 2011, *75*(20): p. 6239-6254.

[213] United States Naval Research Laboratory Center for Computational Materials Science, *Crystal lattice structures web page* 2007, http://cst-www.nrl.navy.mil/lattice/ accessed April 19, 2012.

[214] P. K. Abraitis, R. A. D. Pattrick and D. J. Vaughan, *Variations in the compositional, textural and electrical properties of natural pyrite: A review* International Journal of Mineral Processing 2004, *74*(1-4): p. 41-59.

[215] J. C. Gerard, D. A. Hauglustaine and L. M. Francois, *The faint young sun climatic paradox: A simulation with an interactive seasonal climate-sea ice model* Palaeogeography, Palaeoclimatology, Palaeoecology 1992, *97*(3): p. 133-150.

[216] R. S. Martin, T. A. Mather and D. M. Pyle, *Volcanic emissions and the early earth atmosphere* Geochimica Et Cosmochimica Acta 2007, *71*(15): p. 3673-3685.

[217] J. S. Hudson, J. F. Eberle, R. H. Vachhani, L. C. Rogers, J. H. Wade, R. Krishnamurthy and G. Springsteen, *A unified mechanism for abiotic adenine and purine synthesis in formamide* Angewandte Chemie International Edition 2012, p. 1-5.

[218] D. Cacace, H. Ashbaugh, N. Kouri, S. Bledsoe, S. Lancaster and S. Chalk, *Spectrophotometric determination of aqueous cyanide using a revised phenolphthalin method* Analytica Chimica Acta 2007, *589*(1): p. 137-141.

[219] T. Imanari, S. Tanabe and T. Toida, *Simultaneous determination of cyanide and thiocyanate by high-performance liquid-chromatography* Chemical & Pharmaceutical Bulletin 1982, *30*(10): p. 3800-3802.

[220] J. L. Lambert, J. Ramasamy and J. V. Paukstelis, *Stable reagents for colorimetric determination of cyanide by modified konig reactions* Analytical Chemistry 1975, *47*(6): p. 916-918.

[221] B. A. Logue, D. M. Hinkens, S. I. Baskin and G. A. Rockwood, *The analysis of cyanide and its breakdown products in biological samples* Critical Reviews in Analytical Chemistry 2010, *40*(2): p. 122-147.

[222] J. A. Ma and P. K. Dasgupta, *Recent developments in cyanide detection: A review* Analytica Chimica Acta 2010, *673*(2): p. 117-125.

[223] R. D. Perry and C. L. San Clemente, *Determination of iron with bathophenanthroline following an improved procedure for reduction of iron(iii) ions* Analyst 1977, *102*(1211): p. 114-119.

[224] G. F. Smith, W. H. McCurdy and H. Diehl, *The colorimetric determination of iron in raw and treated municipal water supplies by use of 4 - 7-diphenyl-1 - 10-phenanthroline* Analyst 1952, *77*(917): p. 418-422.

[225] G. K. Mittapalli, K. R. Reddy, H. Xiong, O. Munoz, B. Han, F. De Riccardis, R. Krishnamurthy and A. Eschenmoser, *Mapping the landscape of potentially primordial informational oligomers: Oligodipeptides and oligodipeptoids tagged with triazines as recognition elements* Angewandte Chemie-International Edition 2007, *46*(14): p. 2470-2477.

[226] S. B. Kalidindi and B. R. Jagirdar, *Nanocatalysis and prospects of green chemistry* Chemsuschem 2012, *5*(1): p. 65-75.

[227] M. Yucel, A. Gartman, C. S. Chan and G. W. Luther, *Hydrothermal vents as a kinetically stable source of iron-sulphide-bearing nanoparticles to the ocean* Nature Geoscience 2011, *4*(6): p. 367-371.

[228] M. Halmann, *Cyanamide-induced condensation reactions of glycine* Archives of Biochemistry and Biophysics 1968, *128*(3): p. 808-&.

[229] D. W. Nooner, E. Sherwood, M. A. More and J. Oro, *Cyanamide mediated syntheses under plausible primitive earth conditions .3. Synthesis of peptides* Journal of Molecular Evolution 1977, *10*(3): p. 211-220.

[230] C. F. Chyba, *The hazy details of early earth's atmosphere response* Science 2010, *330*(6005): p. 755-U727.

[231] N. Dauphas and J. F. Kasting, *Low p(co2) in the pore water, not in the archean air* Nature 2011, *474*(7349): p. E2-E3.

[232] C. T. Reinhard and N. J. Planavsky, *Mineralogical constraints on precambrian p(co2)* Nature 2011, *474*(7349): p. E1-E2.

[233] D. Voet and J. G. Voet, *Biochemistry*, John Wiley & Sons, Inc., Hoboken, N.J., 2004.

[234] A. J. Kirby and W. P. Jencks, *Reactivity of nucleophilic reagents toward p-nitrophenyl phosphate dianion* Journal of the American Chemical Society 1965, *87*(14): p. 3209-3216.

[235] P. J. O'Brien and D. Herschlag, *Functional interrelationships in the alkaline phosphatase superfamily: Phosphodiesterase activity of escherichia coli alkaline phosphatase* Biochemistry 2001, *40*(19): p. 5691-5699.

[236] M. A. Pasek, *Rethinking early earth phosphorus geochemistry* Proceedings of the National Academy of Sciences of the United States of America 2008, *105*(3): p. 853-858.

[237] R. L. Frost and M. Weier, *Raman spectroscopic study of vivianites of different origins* Neues Jahrbuch Fur Mineralogie-Monatshefte 2004, 10): p. 445-463.

[238] C. A. McCammon and R. G. Burns, *The oxidation mechanism of vivianite as studied by mossbauer-spectroscopy* American Mineralogist 1980, *65*(3-4): p. 361-366.

[239] J. M. Hollis, F. J. Lovas, A. J. Remijan, P. R. Jewell, V. V. Ilyushin and I. Kleiner, *Detection of acetamide (ch3conh2): The largest interstellar molecule with a peptide bond* Astrophysical Journal 2006, *643*(1): p. L25-L28.

[240] D. Fossé, J. Cernicharo, M. Gerin and P. Cox, *Molecular carbon chains and rings in tmc-1* The Astrophysical Journal 2001, *552*(1): p. 168.

[241] L. E. B. Johansson, C. Andersson, J. Elldér, P. Friberg, A. Hjalmarson, B. Höglund, W. Irvine, H. Olofsson and G. Rydbeck, *Spectral scan of orion a and irc+ 10216 from 72 to 91 ghz* Astronomy & Astrophysics 1984, *130*(p. 227-256.

[242] A. J. Remijan, J. M. Hollis, F. J. Lovas, D. F. Plusquellic and P. R. Jewell, *Interstellar isomers: The importance of bonding energy differences* The Astrophysical Journal 2005, *632*(1): p. 333.

[243] S. Aiello, L. Morbidelli and L. Ulivi, *Far uv radiation transfer and h2co lifetime in dense interstellar clouds* Astrophysics and Space Science 1981, *80*(p. 173-187.

[244] N. Kaifu, K. Takagi and T. Kojima, *Excitation of interstellar methylamine* Astrophysical Journal 1975, *198*(2): p. L85-L88.

[245] J. E. Dickens, W. M. Irvine, M. Ohishi, M. Ikeda, S. Ishikawa, A. Nummelin and A. Hjalmarson, *Detection of interstellar ethylene oxide (c-c2h4o)* Astrophysical Journal 1997, *489*(2): p. 753-757.

[246] L. M. Ziurys, *The chemistry in circumstellar envelopes of evolved stars: Following the origin of the elements to the origin of life* Proceedings of the National Academy of Sciences 2006, *103*(33): p. 12274-12279.

[247] J. M. Hollis, S. N. Vogel, L. E. Snyder, P. R. Jewell and F. J. Lovas, *The spatial scale of glycolaldehyde in the galactic center* Astrophysical Journal 2001, *554*(1): p. L81-L85.

[248] D. M. Mehringer, L. E. Snyder and Y. T. Miao, *Detection and confirmation of interstellar acetic acid* Astrophysical Journal 1997, *480*(1): p. L71-&.

[249] B. Zuckerman, B. E. Turner, D. R. Johnson, F. O. Clark, F. J. Lovas, N. Fourikis, P. Palmer, M. Morris, A. E. Lilley, J. A. Ball, C. A. Gottlieb, M. M. Litvak and H. Penfield, *Detection of interstellar trans-ethyl alcohol* Astrophysical Journal 1975, *196*(3): p. L99-L102.

[250] L. E. Snyder, F. J. Lovas, D. M. Mehringer, N. Y. Miao, Y. J. Kuan, J. M. Hollis and P. R. Jewell, *Confirmation of interstellar acetone* Astrophysical Journal 2002, *578*(1): p. 245-255.

[251] J. M. Hollis, F. J. Lovas, P. R. Jewell and L. H. Coudert, *Interstellar antifreeze: Ethylene glycol* Astrophysical Journal 2002, *571*(1): p. L59-L62.

[252] O. Abbas, C. Rebufa, N. Dupuy and J. Kister, *Ftir-multivariate curve resolution monitoring of photo-fenton degradation of phenolic aqueous solutions comparison with hplc as a reference method* Talanta 2008, *77*(1): p. 200-209.

[253] G. P. Anipsitakis and D. D. Dionysiou, *Transition metal/uv-based advanced oxidation technologies for water decontamination* Applied Catalysis B-Environmental 2004, *54*(3): p. 155-163.

[254] R. Bauer and H. Fallmann, *The photo-fenton oxidation - a cheap and efficient wastewater treatment method* Research on Chemical Intermediates 1997, *23*(4): p. 341-354.

[255] E. Evgenidou, I. Konstantinou, K. Fytianos and I. Poulios, *Oxidation of two organophosphorous insecticides by the photo-assisted fenton reaction* Water Research 2007, *41*(9): p. 2015-2027.

[256] B. L. Bass, *RNA editing - an i for editing* Current Biology 1995, *5*(6): p. 598-600.

[257] U. Nagaswamy, N. Voss, Z. D. Zhang and G. E. Fox, *Database of non-canonical base pairs found in known RNA structures* Nucleic Acids Research 2000, *28*(1): p. 375-376.

[258] J. Stombaugh, C. L. Zirbel, E. Westhof and N. B. Leontis, *Frequency and isostericity of RNA base pairs* Nucleic Acids Research 2009, *37*(7): p. 2294-2312.

[259] T. Inoue and L. E. Orgel, *A non-enzymatic RNA-polymerase model* Science 1983, *219*(4586): p. 859-862.

[260] E. D. Horowitz and N. V. Hud, *Ethidium and proflavine binding to a 2 ',5 '-linked RNA duplex* Journal of the American Chemical Society 2006, *128*(48): p. 15380-15381.

[261] E. D. Horowitz, S. Lilavivat, B. W. Holladay, M. W. Germann and N. V. Hud, *Solution structure and thermodynamics of 2',5' RNA intercalation* Journal of the American Chemical Society 2009, *131*(16): p. 5831-5838.

[262] J. C. Ma and D. A. Dougherty, *The cation,àíœä interaction* Chemical Reviews 1997, *97*(5): p. 1303-1324.

[263] L. S. Lerman, *Structural considerations in interaction of DNA and acridines* Journal of Molecular Biology 1961, *3*(1): p. 18-30.

[264] J. B. Chaires, *Energetics of drug-DNA interactions* Biopolymers 1997, *44*(3): p. 201-215.

[265] K. V. Miroshnychenko and A. V. Shestopalova, *The effect of drug-DNA interactions on the intercalation site formation* International Journal of Quantum Chemistry 2010, *110*(1): p. 161-176.

[266] R. J. H. Davies, *Complexes of polyadenylic acid and 7-methyl-xanthine and their optical-rotatory dispersion* Biochemical and Biophysical Research Communications 1973, *52*(3): p. 1115-1122.

[267] R. J. H. Davies, *Complex-formation between polyadenylic acid and formycin-b* Journal of Molecular Biology 1973, *73*(3): p. 317-327.

[268] R. J. H. Davies, *Complexes of poly(adenylic acid) with complementary monomers* European Journal of Biochemistry 1976, *61*(1): p. 225-236.

[269] P. Giri and G. S. Kumar, *Isoquinoline alkaloids and their binding with polyadenylic acid: Potential basis of therapeutic action* Mini-Reviews in Medicinal Chemistry 2010, *10*(7): p. 568-577.

[270] J. S. Ren and J. B. Chaires, *Sequence and structural selectivity of nucleic acid binding ligands* Biochemistry 1999, *38*(49): p. 16067-16075.

[271] K. S. Huang, M. J. Haddadin, M. M. Olmstead and M. J. Kurth, *Synthesis and reactions of some heterocyclic azacyanines* Journal of Organic Chemistry 2001, *66*(4): p. 1310-1315.

[272] J.-H. Tan, L.-Q. Gu and J.-Y. Wu, *Design of selective g-quadruplex ligands as potential anticancer agents* Mini Reviews In Medicinal Chemistry 2008, *8*(11): p. 1163-1178.

[273] H. M. Berman, W. Stallings, H. L. Carrell, J. P. Glusker, S. Neidle, G. Taylor and A. Achari, *Molecular and crystal-structure of an intercalation complex - proflavine-cytidylyl-(3',5')-guanosine* Biopolymers 1979, *18*(10): p. 2405-2429.

[274] T. Schelhorn, S. Kretz and H. W. Zimmermann, *Reinvestigation of the binding of proflavine to DNA - is intercalation the dominant binding effect* Cellular and Molecular Biology 1992, *38*(4): p. 345-365.

[275] P. Stevenson, K. R. Sones, M. M. Gicheru and E. K. Mwangi, *Comparison of isometamidium chloride and homidium bromide as prophylactic drugs for trypanosomiasis in cattle at nguruman, kenya* Acta Tropica 1995, *59*(2): p. 77-84.

[276] D. P. Heller and C. L. Greenstock, *Fluorescence lifetime analysis of DNA intercalated ethidium-bromide and quenching by free dye* Biophysical Chemistry 1994, *50*(3): p. 305-312.

[277] J. Olmsted and D. R. Kearns, *Mechanism of ethidium-bromide fluorescence enhancement on binding to nucleic-acids* Biochemistry 1977, *16*(16): p. 3647-3654.

[278] J. B. Le Pecq and C. Paoletti, *A fluorescent complex between ethidium bromide and nucleic acids - physical-chemical characterization* Journal of Molecular Biology 1967, *27*(1): p. 87-106.

[279] A. Canals, M. Purciolas, J. Aymani and M. Coll, *The anticancer agent ellipticine unwinds DNA by intercalative binding in an orientation parallel to base pairs* Acta Crystallographica Section D-Biological Crystallography 2005, *61*(p. 1009-1012.

[280] K. W. Kohn, M. J. Waring, D. Glaubiger and C. A. Friedman, *Intercalative binding of ellipticine to DNA* Cancer Research 1975, *35*(1): p. 71-76.

[281] N. C. Garbett and D. E. Graves, *Extending nature's leads: The anticancer agent ellipticine* Current Medicinal Chemistry - Anti-Cancer Agents 2004, *4*(2): p. 149-172.

[282] C. Auclair, *Multimodal action of antitumor agents on DNA - the ellipticine series* Archives of Biochemistry and Biophysics 1987, *259*(1): p. 1-14.

[283] S. Pal, G. S. Kumar, D. Debnath and M. Maiti, *Interaction of the antitumour alkaloid coralyne with duplex deoxyribonucleic acid structures: Spectroscopic and viscometric studies* Indian Journal of Biochemistry & Biophysics 1998, *35*(6): p. 321-332.

[284] F. F. Xing, G. T. Song, J. S. Ren, J. B. Chaires and X. G. Qu, *Molecular recognition of nucleic acids: Coralyne binds strongly to poly(a)* Febs Letters 2005, *579*(22): p. 5035-5039.

[285] M. Polak and N. V. Hud, *Complete disproportionation of duplex poly(dt)center dot poly(da) into triplex poly(dt)center dot poly(da)center dot poly(dt) and poly(da) by coralyne* Nucleic Acids Research 2002, *30*(4): p. 983-992.

[286] O. P. Cetinkol, A. E. Engelhart, R. K. Nanjunda, W. D. Wilson and N. V. Hud, *Submicromolar, selective g-quadruplex ligands from one pot: Thermodynamic and structural studies of human telomeric DNA binding by azacyanines* Chembiochem 2008, *9*(12): p. 1889-1892.

[287] J. B. Chaires, K. R. Fox, J. E. Herrera, M. Britt and M. J. Waring, *Site and sequence specificity of the daunomycin DNA interaction* Biochemistry 1987, *26*(25): p. 8227-8236.

[288] W. Muller and D. M. Crothers, *Studies of binding of actinomycin and related compounds to DNA* Journal of Molecular Biology 1968, *35*(2): p. 251-290.

[289] K. E. van Holde, W. C. Johnson and P. S. Ho, *Principles of physical biochemistry*, Pearson Education, Inc., Upper Saddle River, NJ, 2006.

[290] W. Likussar, *Computer approach to continuous variations method for spectrophotometric determination of extraction and formation constants* Analytical Chemistry 1973, *45*(11): p. 1926-1931.

[291] W. Likussar and D. F. Boltz, *Theory of continuous variations plots and a new method for spectrophotometric determination of extraction and formation constants* Analytical Chemistry 1971, *43*(10): p. 1265-1272.

[292] D. A. Skoog, D. M. West, F. J. Holler and S. R. Crouch, *Fundamentals of analytical chemistry*, Brooks/Cole - Thomson Learning, Belmont, CA, 2004.

[293] X. G. Qu and J. B. Chaires *Analysis of drug-DNA binding data*, in Methods in Enzymology: Numerical Computer Methods, Part C, 2000, *321*: pp. 353-369.

[294] M. Eriksson and B. Norden *Linear and circular dichroism of drug-nucleic acid complexes*, in Methods in Enzymology: Drug-Nucleic Acid Interactions, 2001, *340*: pp. 68-98.

[295] K. J. Laidler, J. H. Meiser and B. C. Sanctuary, *Physical chemistry*, Houghton Mifflin Company, Boston, 2003.

[296] D. Case, T. Darden, I. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, K. Merz, B. Wang, D. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J. Caldwell, W. Ross and P. A. Kollman *AMBER 11*, University of California, San Francisco (USA), 2011.

[297] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, *The r.E.D. Tools: Advances in resp and esp charge derivation and force field library building* Physical Chemistry Chemical Physics 2010, *12*(28): p. 7821-7839.

[298] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute and K. Zakrzewska, *Conformational analysis of nucleic acids revisited: Curves* Nucleic Acids Research 2009, *37*(17): p. 5917-5929.

[299] V. Bloomfield, D. M. Crothers and I. Tinoco, *Nucleic acids: Structure, properties, and function*, University Science Books, 2000.

[300] C. Roberts, R. Bandaru and C. Switzer, *Theoretical and experimental study of isoguanine and isocytosine: Base pairing in an expanded genetic system* Journal of the American Chemical Society 1997, *119*(20): p. 4640-4649.

[301] F. Seela, C. F. Wei and Z. Kazimierczuk, *Substituent reactivity and tautomerism of isoguanosine and related nucleosides* Helvetica Chimica Acta 1995, *78*(7): p. 1843-1854.

[302] J. Sepiol, Z. Kazimierczuk and D. Shugar, *Tautomerism of iso-guanosine and solvent-induced keto-enol equilibrium* Zeitschrift Fur Naturforschung C-a Journal of Biosciences 1976, *31*(7-8): p. 361-370.

[303] J. L. Mergny and L. Lacroix, *Analysis of thermal melting curves* Oligonucleotides 2003, *13*(6): p. 515-537.

[304] F. Quadrifoglio and Crescenz.V, *Binding of actinomycin and of daunomycin to DNA - calorimetric and spectroscopic investigation* Biophysical Chemistry 1974, *2*(1): p. 64-69.

[305] C. R. Geyer, T. R. Battersby and S. A. Benner, *Nucleobase pairing in watson-crick-like genetic expanded information systems* Structure 2003, *11*(12): p. 1485-1498.

[306] Z. H. Sun and L. W. McLaughlin, *Effects of the minor groove pyrimidine nucleobase functional groups on the stability of duplex DNA: The impact of uncompensated minor groove amino groups* Biopolymers 2007, *87*(2-3): p. 183-195.

[307] J. T. Goodwin and D. G. Lynn, *Template-directed synthesis - use of a reversible-reaction* Journal of the American Chemical Society 1992, *114*(23): p. 9197-9198.